Myke King

Process Control

A Practical Approach

Second Edition

with website

WILEY

Process Control
A Practical Approach

Process Control
A Practical Approach

Second Edition

Myke King
Whitehouse Consulting, Isle of Wight, UK

WILEY

This edition first published 2016

First Edition published in 2011

Registered Office
John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Library of Congress Cataloging-in-Publication Data

Names: King, Myke, 1951– author.
Title: Process control : a practical approach / Myke King.
Description: Second edition. | Chichester, West Sussex, United Kingdom : John Wiley & Sons Inc., 2016. |
Includes bibliographical references and index.
Identifiers: LCCN 2015044557 (print) | LCCN 2015045760 (ebook) | ISBN 9781119157748 (print) | ISBN 9781119157755 (epdf) |
ISBN 9781119157762 (epub) | ISBN 9781119157779 (Obook) | ISBN 9781119157755 (Adobe PDF) | ISBN 9781119157762 (ePub)
Subjects: LCSH: Chemical process control.
Classification: LCC TP155.75 .K56 2016 (print) | LCC TP155.75 (ebook) | DDC 660/.2815–dc23
LC record available at http://lccn.loc.gov/2015044557

A catalogue record for this book is available from the British Library.

Cover image: gettyimages/imagewerkz

Set in 10/12pt Times by SPi Global, Pondicherry, India

1 2016

Contents

Preface

So why write yet another book on process control? There are already many published. But they are largely written by academics and intended mainly to support courses taught at universities. Excellent as some of these books are in meeting that aim, the content of many academic courses has only limited relevance to control design in the process industry. There are a few books that take a more practical approach but these usually provide only an introduction to the technologies. They contain enough detail if used as part of a wider engineering course but not enough for the practitioner. This book aims more to meet the needs of industry.

So why write a second edition? Most of the process control techniques covered in the first edition have been in use for decades and will continue to be so for decades to come. While there are ongoing developments in the hardware of control, there is little drive to move away from well-established techniques. What prompted the second edition was the seemingly never-ending desire of control engineers to better understand the subject and to explore new ways of applying these techniques. Almost none of the material in the first edition has been deleted; most of it has been expanded with improved explanations and, where valuable, greater depth.

Most engineers responsible for the design and maintenance of control applications find daunting much of the theoretical mathematics that is common in the academic world. In this book we have aimed to keep the mathematics to a minimum. For example, Laplace transforms are only included so that the reader may relate what is in this book to what will be found in most theoretical texts and in the documentation provided by many DCS (distributed control system) vendors. They are not used in any of the control design techniques. And while we present the mathematical derivation of these techniques, to show that they have a sound engineering basis, the reader can skip these if too daunting and simply apply the end result.

This edition has two substantial new chapters. These were added with some trepidation. They appear to be in conflict with the objective of minimising the use of more advanced mathematics. Indeed they have doubled the total number of equations in the book. While the reader may be unfamiliar with some of the mathematical symbols, supporting explanations have been kept simple as possible. These, together with the examples of practical application, should help the more determined readily apply the methods. Those less enthusiastic can however skip these chapters and still be very competent control engineers.

The book aims to present techniques that have an immediate practical application. In addition to the design methods, it describes any shortcuts that can be taken and how to avoid common pitfalls. The methods have been applied on many processes to a wide range of controllers. They should work.

In addition to providing effective design methods, this book should improve the working practices of many control engineers. For example, the majority still prefer to tune PID (proportional, integral, derivative) controllers by trial-and-error. This is time-consuming and rarely leads to controllers performing as well as they should. This might be because of a justified mistrust of published tuning methods. Most do have serious limitations. This book addresses this and offers a method proven to be effective in terms of both controller performance and engineering effort.

DCS include a wide array of control algorithms with many additional engineer-definable parameters. The DCS vendors are poor at explaining the purpose of these algorithms with the result that the industry is rife with misinterpretation of their advantages and disadvantages. These algorithms were included in the

original system specification by engineers who knew their value. But this knowledge has not passed to the industry. The result is that there are substantial improvements that can be made on almost every process unit, surpassing what the control engineer is even aware of – let alone know how to implement. This book addresses all the common enhancements.

This book takes a back-to-basics approach. The use of MPC (multivariable process control, or multivariable predictive control) is widespread in industry. Control engineering staff and their contractors have invested thousands of man-hours in the necessary plant testing and commissioning. Improving the basic controls is not usually an option once the MPC is in place. Improvements are likely to change the process dynamics and would thus involve substantial re-engineering of the MPC. Thus poor basic control remains the status quo and becomes the accepted standard to the point where it is not addressed even when the opportunity presents itself. This book raises the standard of what might be expected from the performance of basic controls.

Before MPC, ARC (advanced regulatory control) was commonplace. MPC has rightly replaced many of the more complex ARC techniques. But it has been used by too many as the panacea to any control problem. There remain many applications where ARC outperforms MPC; but appreciation of its advantages is now hard to find in industry. The expertise to apply it is even rarer. This book aims to get the engineer to reconsider where ARC should be applied and to help develop the necessary implementation skills.

However, due credit must be given to MPC as a major step forward in the development of APC (advanced process control) techniques. This book focuses on how to get the best out of its application, rather than replicate the technical details that appear in many text books, papers and product documentation.

The layout of the book has been designed so that the reader can progress from relatively straightforward concepts through to more complex techniques applied to more complex processes. It is assumed that the new reader is comfortable with mathematics up to a little beyond high school level. As the techniques become more specific some basic knowledge of the process is assumed, but introductory information is included – particularly where it is important to control design. Heavily mathematical material, daunting to novices and not essential to successful implementation, has been relegated to the end of the book.

SI units have been mainly used throughout but, where important and practical, conversion to imperial units is given in the text. Methods published in non-SI units have been included without change if doing so would make them too complex.

The book is targeted primarily for use in the continuous process industry. But even predominantly batch plants have continuous controllers and often have sections of the process which are continuous. My experience is mainly in the oil and petrochemicals industries and, despite every effort being taken to make the process examples as generic as possible, it is inevitable that this will show through. However, this should not be seen as a reason for not applying the techniques in other industries. Many started there and have been applied by others to a wide range of processes.

Academic institutions are beginning to appreciate the need to make their courses more relevant to the process industry. These institutions are traditionally ranked according to degree results and research work. Now they are becoming increasingly dependent on being selected by students based on their reputation for supplying industry with appropriately educated engineers. While there has been some progress, practitioners still perceive a huge gulf between theory and practice. Of course there is a place for the theory. Many of the modern control technologies now applied in the process industry are developed from it. And there are other industries, such as aerospace, where it is essential.

The debate is what should be taught as part of chemical engineering. Very few chemical engineers benefit from the theory traditionally included. Indeed the risk is that many potentially excellent control engineers do not enter the profession because of the poor image that theoretical courses create. Further, those that do follow a career in process control, can find themselves working in an organisation managed by a chemical engineering graduate who has no appreciation of what process control technology can do and its

importance to the business. The challenge for the academic world is to not only recognise the need to change course content but to find lecturers that have extensive industrial experience. While not a substitute for experience, this book goes a long way towards addressing this need and indeed has been adopted by several university chemical engineering departments.

It is the nature of almost any engineering subject that the real gems of useful information get buried in amongst the background detail. Listed here are the main items worthy of special attention by the engineer because of the impact they can have on the effectiveness of control design.

- Understanding the process dynamics is essential to the success of almost every process control technique. These days there is very little excuse for not obtaining these by plant testing or from historically collected data. There are a wide range of model identification products available plus enough information is given in Chapter 2 for a competent engineer to develop a simple spreadsheet-based application.

- Often overlooked is the impact that apparently unrelated controllers can have on process dynamics. Their tuning and whether they are in service or not, will affect the result of step-tests and hence the design of the controller. Any changes made later can then severely disrupt controller performance. How to identify such controllers, and how to handle their effect, are described in Chapters 2 and 8.

- Modern DCS include a number of versions of the PID controller. Of particular importance in the proportional-on-PV algorithm. It is probably the most misunderstood option and is frequently dismissed as too slow compared to the more conventional proportional-on-error version. In fact, if properly tuned, it can make a substantial improvement to the way that process disturbances are dealt with – often shortening threefold the time it takes the process to recover. This is fully explained in Chapter 3.

- Controller tuning by trial-and-error should be seen as an admission of failure to follow proper design procedures, rather than the first choice of technique. To be fair to the engineer, every published tuning technique and most proprietary packages have serious limitations. Chapter 3 presents a new technique that is well-proven in industry and gives sufficient information for the engineer to extend it as required to accommodate special circumstances.

- Derivative action is too often excluded from controllers. Understandably introducing a third parameter to tune by trial-and-error might seem an unnecessary addition to workload. It also has a poor reputation in the way that it amplifies measurement noise. But, engineered using the methods in Chapter 3, it can substantially lessen the impact of process disturbances.

- Tuning level controllers to exploit surge capacity can dramatically improve the stability of the process. However, the ability to achieve this is too frequently restricted by poor instrument design. And, often it is not implemented because of difficulty in convincing the plant operator that the level should be allowed to deviate from SP (set-point) for long periods. Chapter 4 describes the important aspects in sizing and locating the level transmitter and how the conventional linear PID algorithm can be tuned – without the need even to perform any plant testing. It also shows how nonlinear algorithms, particularly gap control, can be set up to handle the situation where the size of the flow disturbances can vary greatly.

- While many will appreciate how signal conditioning can be applied to measurements and controller outputs to help linearise the behaviour, not so commonly understood is how it can be applied to constraint controllers. Doing so can enable constraints to be approached more closely and any violation dealt with more quickly. Full details are given in Chapter 5.

- Many engineers are guilty of installing excessive filtering to deal with noisy measurements. Often implemented only to make trends look better they introduce additional lag and can have a detrimental impact on controller performance. Chapter 5 gives guidance on when to install a filter and offers a new type that actually reduces the overall process lag.

- Split-ranging is commonly used to allow two or more valves to be moved sequentially by the same controller. While successful in some cases, the technique is prone to problems with linearity and discontinuity. A more reliable alternative is offered in Chapter 5.

- Feedforward control is often undervalued or left to the MPC. Chapter 6 shows how simple techniques, applied to few key variables, can improve process stability far more effectively than MPC.

- A commonly accepted problem with MPC is that, if not properly monitored, it becomes over-constrained. In fact, if completely neglected, they are effectively fully disabled – even though they may show 100% up-time. Chapter 8 offers a range of monitoring tools, supplementary to those provided by the MPC vendor, which can be readily configured by the engineer.

- There are many examples of MPC better achieving the wrong operating objective; unbeknown to the implementer they are reducing process profitability. Rather than attempt to base the cost coefficients on real economics they are often adjusted to force the MPC to follow the historically accepted operating strategy. Some MPC schemes are extremely complex and it is unlikely that even the most competent plant manager will have considered every opportunity for adopting a different strategy. Chapter 12 shows how properly setting up the MPC can reveal such opportunities.

- There are literally thousands of inferential properties, so-called 'soft sensors', in use today that are ineffective. Indeed many of them are so inaccurate that process profitability would be improved by decommissioning them. Chapter 9 shows how many of the statistical techniques that are used to assess their accuracy are flawed and can lead the engineer into believing that their performance is adequate. It also demonstrates that automatically updating the inferential bias with laboratory results will generally aggravate the problem.

- Simple monitoring of on-stream analysers, described in Chapter 9, ensures that measurement failure does not disrupt the process and that the associated reporting tools can do much to improve their reliability and use.

- Compensating fuel gas flow measurement for variations in pressure, temperature and molecular weight requires careful attention. Done for accounting purposes, it can seriously degrade the performance of fired heater and boiler control schemes. Chapter 10 presents full details on how it should be done.

- Manipulating fired heater and boiler duty by control of fuel pressure, rather than fuel flow, is common practice. However, it restricts what improvements can be made to the controller to better handle process disturbances. Chapter 10 shows how the benefits of both approaches can be captured.

- Fired heater pass balancing is often installed to equalise pass temperatures in order to improve efficiency. Chapter 10 shows that the fuel saving is negligible and that, in some cases, the balancing may accelerate coking. However, there may be much larger benefits available from the potential to debottleneck the heater.

- Compressor control packages are often supplied as 'black boxes' and many compressor manufacturers insist on them being installed in special control systems on the basis that DCS-based schemes would be too slow. Chapter 11 describes how these schemes work and, using the tuning method in Chapter 3, how they might be implemented in the DCS.

- A common failing in many distillation column control strategies is the way in which they cope with changes in feed rate and composition. Often only either the reboiler duty or the reflux flow is adjusted to compensate – usually under tray temperature control. Chapter 12 shows that failing to adjust both is worse than making no compensation. Other common misconceptions include the belief that column pressure should always be minimised and that the most economic strategy is always to exactly meet all product specifications.

- There is a growing number of products offered to monitor basic controllers. Many report parameters not fully understood by many engineers and, worse, do not properly measure performance. Many of the techniques are the product of naïve academic research and their practicality has yet to been demonstrated. Chapter 13 describes commonly used methods – the more valuable of which can readily be implemented by the control engineer without incurring costly software licence fees. Chapters 14 and 15 give more detail of the mathematics behind the techniques.

- There are many pitfalls in executing an advanced control project. Significant profit improvement opportunities are often overlooked because of the decision to go with a single supplier for the benefits study, MPC package, inferentials and implementation. Basic controls, inferentials and advanced regulatory controls are not given sufficient attention before awarding the implementation contract. The need for long-term application support is often under-estimated and poor management commitment will jeopardise the capture of benefits. Chapter 13 describes how these and many other issues can be addressed.

- Statistical methods are frequently used by control engineers for benefit estimation, reconciling process data, the development of inferentials, dynamic model identification and for monitoring the performance of process control. Like much of control theory, statistical techniques are often described in a highly mathematical style. Chapter 14 aims to present, in a way which makes them more readily understandable, those methods potentially valuable to the control engineer.

- Process control texts are often criticised for containing excessive and complex mathematics. Indeed one of the aims of this book was not to make the same mistake. However, there are many techniques, perhaps not originally intended for use in process control, which merit attention by the control engineer. Chapter 15 provides a more in-depth understanding of the mathematical techniques used throughout the book and also describes others worthy of consideration.

Gaining the knowledge and experience now contained in this book would have been impossible if it were not for the enthusiasm and cooperation of my clients. I am exceedingly grateful to them. In particular I must mention the support provided by Ed Dilley who read not only every word of the first edition but also the draft of this edition. His suggestions were invaluable in improving both accuracy and understanding.

Myke King
Isle of Wight

About the Author

Myke King is the founder and director of Whitehouse Consulting, an independent consulting organisation specialising in process control. He has over 40 years' experience working with over 100 clients from more than 30 countries. As part of his consulting activities Myke has developed training courses covering all aspects of process control. To date, around 2,000 delegates have attended these courses. He also lectures at several universities. To support his consulting activities he has developed a range of software to streamline the design of controllers and to simulate their use for learning exercises.

Myke graduated from Cambridge University in the UK with a Master's degree in chemical engineering. His course included process control taught as part of both mechanical engineering and chemical engineering. At the time he understood neither. On graduating he joined, by chance, the process control section at Exxon's refinery at Fawley in the UK. Fortunately he quickly discovered that the practical application of process control bore little resemblance to the theory he had covered at university. He later became head of the process control section and then moved to operations department as a plant manager. This was followed by a short period running the IT section.

Myke left Exxon to co-found KBC Process Automation, a subsidiary of KBC Process Technology, later becoming its managing director. The company was sold to Honeywell where it became their European centre of excellence for process control. It was at this time Myke set up Whitehouse Consulting.

Myke is a Fellow of the Institute of Chemical Engineers in the UK.

1

Introduction

In common with many introductions to the subject, process control is described here in terms of layers. At the lowest level is the process itself. Understanding the process is fundamental to good control design. While the control engineer does not need the level of knowledge of a process designer, an appreciation of how the process works, its key operating objectives and basic economics is vital. In one crucial area the control engineer's knowledge must exceed that of the process engineer, who needs primarily an understanding of the *steady-state* behaviour. The control engineer must also understand the *process dynamics*, i.e. how process parameters move between steady states.

Next up is the field instrumentation layer, comprising measurement transmitters, control valves and other actuators. This layer is the domain of instrument engineers and technicians. However, control engineers need an appreciation of some of the hardware involved in control. They should to be able to recognise a measurement problem or a control valve working incorrectly. They must be aware of the accuracy, linearity and the dynamic behaviour of instrumentation - and understand how these issues should be dealt with.

Above the field instrumentation is the DCS and process computer. These will be supported by a system engineer. It is normally the control engineers' responsibility to configure the control applications, and their supporting graphics, in the DCS. So they need to be well-trained in this area. In some sites, only the system engineer is permitted to make changes to the system. However, this does not mean that the control engineer does not need a detailed understanding of how it is done. Close cooperation between control engineer and system engineer is essential.

The lowest layer of process control applications is described as *regulatory control*. This includes all the basic controllers for flow, temperature, pressure and level, but it also includes control of product quality. Regulatory is not synonymous with basic. Regulatory controls are those which maintain the process at a desired condition, or *set-point (SP),* but that does not mean they are simple. They can involve complex instrumentation such as on-stream analysers. They can employ 'advanced' techniques such as signal conditioning, feedforward, dynamic compensation, overrides, inferential properties, etc. Such techniques are often described as *advanced regulatory control (ARC)*. Generally they are implemented within the DCS block structure, with perhaps some custom code, and are therefore sometimes called 'traditional' advanced control. This is the domain of the control engineer.

There will be somewhere a division of what falls into the responsibilities between the control engineer and others working on the instrumentation and system. The simplistic approach is to assign all hardware

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

to these staff and all configuration work to the control engineer. But areas such as algorithm selection and controller tuning need a more flexible approach. Many basic controllers, providing the tuning is reasonable, do not justify particular attention. Work on those that do requires the skill more associated with a control engineer. Sites that assign all tuning to the instrument department risk overlooking important opportunities to improve process performance.

Moving up the hierarchy, the next level is *constraint control*. This comprises control strategies that drive the process towards operating limits, where closer approach to these limits is known to be profitable. Indeed, on continuous processes, this level typically captures the large majority of the available process control benefits. The main technology applied here is *multivariable predictive control (MPC)*. Because of its relative ease of use and its potential impact on profitability, it has become the focus of what is generally known as *advanced process control (APC)*. In fact, as a result, basic control and ARC have become somewhat neglected. Many sites (and many APC vendors) no longer have personnel who appreciate the value of these technologies or have the know-how to implement them.

The topmost layer, in terms of closed loop applications, is *optimisation*. This is based on key economic information such as feed price and availability, product prices and demand, energy costs, etc. Optimisation means different things to different people. The planning group would claim they optimise the process, as would a process support engineer determining the best operating conditions. MPC includes some limited optimisation capabilities. It supports objective coefficients which can be set up to be consistent with process economics. Changing the coefficients can cause the controller to adopt a different strategy in terms of which constraints it approaches. However, those MPC packages based on linear process models cannot identify an unconstrained optimum. This requires a higher fidelity process representation, possibly a rigorous simulation. This we describe as *closed-loop real-time optimisation (CLRTO)* or usually just *RTO*.

Implementation should begin at the base of the hierarchy and work up. Any problems with process equipment or instrumentation will affect the ability of the control applications to work properly. MPC performance will be restricted and RTO usually needs to work in conjunction with the MPC. While all this may be obvious, it is not necessarily reflected in the approach that some sites have towards process control. There are sites investing heavily in MPC but giving low priority to maintaining basic instrumentation. And most give only cursory attention to regulatory control before embarking on implementation of MPC.

2

Process Dynamics

Understanding process dynamics is essential to effective control design. Indeed, as will become apparent in later chapters, most design involves performing simple calculations based solely on a few dynamic parameters. While control engineers will commit several weeks of round-the-clock effort to obtaining the process dynamics for MPC packages, most will take a much less analytical approach to regulatory controls. This chapter aims to demonstrate that process dynamics can be identified easily and that, when combined with the design techniques described in later chapters, will result in controllers that perform well without the need for time-consuming tuning by trial-and-error.

2.1 Definition

To explore dynamic behaviour, as an example, we will use a simple fired heater as shown in Figure 2.1. It has no automatic controls in place and the minimum of instrumentation – a temperature indicator (TI) and a fuel control valve. The aim is to ultimately commission a temperature controller which will use the temperature as its *process variable (PV)* and the fuel valve position as its *manipulated variable (MV)*.

Figure 2.2 shows the effect of manually increasing the opening of the valve. While the temperature clearly rises as the valve is opened, the temperature trend is somewhat different from that of the valve. We use a number of parameters to quantify these differences.

The test was begun with the process steady and sufficient time was given for the process to reach a new steady state. We observe that the steady state change in temperature is different from that of the valve. This difference is quantified by the *steady state process gain* and is defined by the expression

$$\text{process gain} = \frac{\text{change in temperature}}{\text{change in valve position}} \tag{2.1}$$

Process gain, occasionally also called *process sensitivity*, is given the symbol K_p. If we are designing controls to be installed in the DCS, as opposed to a computer-based MPC, K_p should generally have no

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

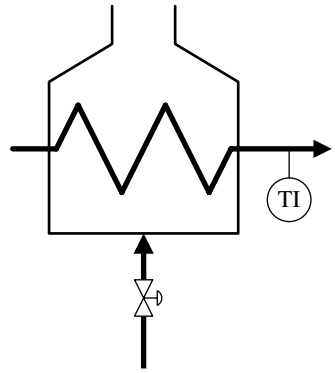

Figure 2.1 *Process diagram*

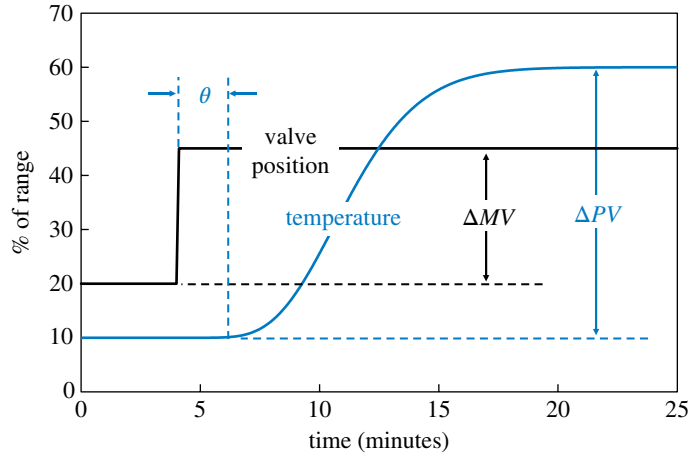

Figure 2.2 *Process response*

dimensions. This is because the DCS works internally with measurements represented as fractions (or percentages) of instrument range.

$$K_p = \frac{\Delta PV}{\Delta MV} \tag{2.2}$$

where

$$\Delta PV = \frac{\text{change in temperature}}{\text{range of temperature transmitter}} \tag{2.3}$$

and

$$\Delta MV = \frac{\text{change in valve position}}{\text{range of valve positioner}} \tag{2.4}$$

Instrument ranges are defined when the system is first configured and generally remain constant. However, it is often overlooked that the process gain changes if an instrument is later re-ranged and, if that instrument is either a PV or MV of a controller, then the controller should be re-tuned to retain the same performance.

Numerically K_p may be positive or negative. In our example, temperature rises as the valve is opened. If we were to increase heater feed rate (and keep fuel rate constant), then the temperature would fall. K_p, with respect to changes in feed rate, would therefore be negative. Nor is there is any constraint on the absolute value of K_p. Very large and very small values are common. In unusual circumstances K_p may be zero; there will be a *transient* disturbance to the PV but it will return to its starting point.

The other differences, in Figure 2.2, between the trends of temperature and valve position are to do with timing. We can see that the temperature begins moving some time after the valve is opened. This delay is known as the *process deadtime*; until we develop a better definition, it is the time difference between the change in MV and the first perceptible change in PV. It is usually given the symbol θ. Deadtime is caused by *transport delays*. Indeed, in some texts, it is described as *transport lag* or *distance velocity lag*. In our example the prime cause of the delay is the time it takes for the heated fluid to move from the firebox to the temperature instrument. We describe later how deadtime can also be introduced by the instrumentation. Clearly the value of θ must be positive but otherwise there is no constraint on its value. Many processes will exhibit virtually no delay; there are some where the delay can be measured in hours or even in days.

Finally, the shape of the temperature trend is very different from that of the valve position. This is caused by the 'inertia' or *capacitance* of the system to store mass or energy. The heater coil will comprise a large mass of steel. Burning more fuel will cause the temperature in the firebox to rise quickly and hence raise the temperature of the external surface of the steel. But it will take longer for this to have an impact on the internal surface of the steel in contact with the fluid. Similarly the coil will contain a large quantity of fluid and it will take time for the bulk temperature to increase. The field instrumentation can add to the lag. For example, the temperature is likely to be a thermocouple located in a steel *thermowell*. The thermowell may have thick walls which cause a lag in the detection of an increase in temperature. Lag is quite different from deadtime. Lag does not delay the start of the change in PV. Without deadtime the PV will begin changing immediately but, because of lag, takes time to reach a new steady state. We normally use the symbol τ to represent lag.

To help distinguish between deadtime and lag, consider liquid flowing at a constant rate (F) into a vessel of volume (V). The process is at steady state. The fraction (x) of a component in the incoming liquid is changed at time zero ($t = 0$) from x_{start} to x_{new}. By mass balance the change in the quantity of the component ($V.dx$) in the vessel is the difference between what has entered less what has left during the time interval (dt). Assuming the liquid is perfectly mixed then, if x is the current fraction in the vessel:

$$V.dx = F.dt.x_{new} - F.dt.x \tag{2.5}$$

Rearranging:

$$\frac{V}{F}\frac{dx}{dt} + x = x_{new} \tag{2.6}$$

The general solution to this equation is:

$$x = A\left(1 - e^{-t/\tau}\right) + B \qquad \text{where} \qquad \tau = \frac{V}{F} \tag{2.7}$$

At the start ($t = 0$) we know that $x = x_{start}$ and at steady state ($t \to \infty$) we know that $x \to x_{new}$; so, solving for A and B, Equation (2.7) becomes:

$$x = \left(x_{new} - x_{start}\right)\left(1 - e^{-t/\tau}\right) + x_{start} \tag{2.8}$$

To simplify, we take the case when $x_{start} = 0$; then:

$$x = x_{new}\left(1 - e^{-t/\tau}\right) \tag{2.9}$$

In the well-mixed case the delay (θ) would be zero. The outlet composition would begin changing immediately, with a lag determined by *V/F* – the *residence time* of the vessel. However, if the vessel was replaced with pipework of the same volume, we could assume no mixing takes place and the change in composition would pass through as a step change delayed by the residence time.

$$\theta = \frac{V}{F} \tag{2.10}$$

In this case the lag would be zero. In practice, neither perfect mixing nor no mixing is likely and the process will exhibit a combination of deadtime and lag.

The DCS will also be a source of deadtime, on average equal to half the controller *sampling period* – more usually known as the *scan interval* (*ts*). For example, if a measurement is scanned every two seconds, there will be a delay of up two seconds in detecting a change. While this is usually insignificant compared to any delay in the process, it is a factor in the design of controllers operating on processes with very fast dynamics – such as compressors. The delay can be increased further by the *resolution* (also called *quantisation*) of the field instrumentation. Resolution is the least interval between two adjacent discrete values that can be distinguished from one another. Imagine that this is 0.1% of range and that the measurement is ramped up 10% of range over an hour. The instrumentation will not report a change until it has exceeded 0.1%; this will incur additional deadtime of 36 seconds. Again, only when the process dynamics are extremely fast, do we have to concern ourselves about the variable delay this can cause. A much larger source of deadtime is discontinuous measurement. This is common for many types of on-stream analysers, such as chromatographs, which take a sample, analyse it and report the result some minutes later. Added to this are delays which might occur in transporting the sample to the analyser and any time required by the analyser preparing to receive the next sample. Such delays are often comparable to the process dynamics and need to be taken account of in controller design.

When trying to characterise the shape of the PV trend we also have to consider the *order (n)* of the process. While, in theory, processes can have very high orders, in practice, we can usually assume that they are first order. However, there are occasions where this assumption can cause problems, so it is important to understand how to recognise this situation.

Conceptually order can be thought of as the number of sources of lag. Figure 2.3 shows a process contrived to demonstrate the effect of combining two lags. It comprises two identical vessels, both open to the atmosphere and both draining through identical valves. Both valves are simultaneously opened fully. The flow through each valve is determined by the head of liquid in the vessel so, as this falls, the flow through the valve reduces and the level falls more slowly.

We will use A as the cross-sectional area of the vessel and h as the height of liquid (starting at 100%). If we assume for simplicity that flow is related linearly to h with k as the constant of proportionality, then

$$A\frac{dh}{dt} = -kh \tag{2.11}$$

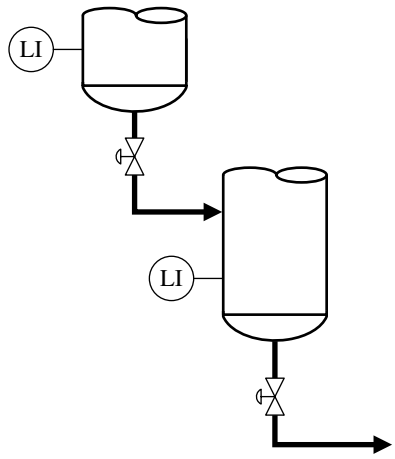

Figure 2.3 *Illustration of order*

Thus

$$A\int_{100}^{h}\frac{dh}{h}=-k\int_{0}^{t}dt \tag{2.12}$$

Integrating gives

$$A\left[\ln(h)\right]_{100}^{h}=-k\left[t\right]_{0}^{t} \tag{2.13}$$

$$h=100e^{-kt/A} \tag{2.14}$$

$$h=100e^{-t/\tau} \qquad \text{where} \quad \tau=\frac{A}{k} \tag{2.15}$$

The shape of the resulting trend is governed by Equation (2.15). Trend A in Figure 2.4 shows the level in the upper vessel. It shows the characteristic of a first order response in that the rate of change of PV is greatest at the start of the change. Trend B shows the level in the lower vessel – a second order process. Since this vessel is receiving liquid from the first then, immediately after the valves are opened, the inlet and outlet flows are equal. The level therefore does not change immediately. This *apparent deadtime* is a characteristic of higher order systems and is additive to any real deadtime caused by transport delays and the instrumentation. Thus by introducing additional deadtime we can approximate a high order process to first order. This approximation is shown as the dashed line.

The accuracy of the approximation is dependent on the combination of process lags. While trend B was drawn with both vessels identical, trend C arises if we increase the lag for the top vessel (e.g. by reducing the size of the valve). We know that the system is still second order but visually the trend could be first order. Our approximation will therefore be very accurate. However, if we reduce the lag of the top vessel below that of the bottom one, then we obtain trend D. This arises because, on opening both valves, the flow entering the bottom vessel is greater than that leaving and so the level initially rises. This is *inverse response*; the PV initially moves in a direction opposite to the steady-state change. Fitting a first order model to this response would be extremely inaccurate. Examples of processes prone to this type of

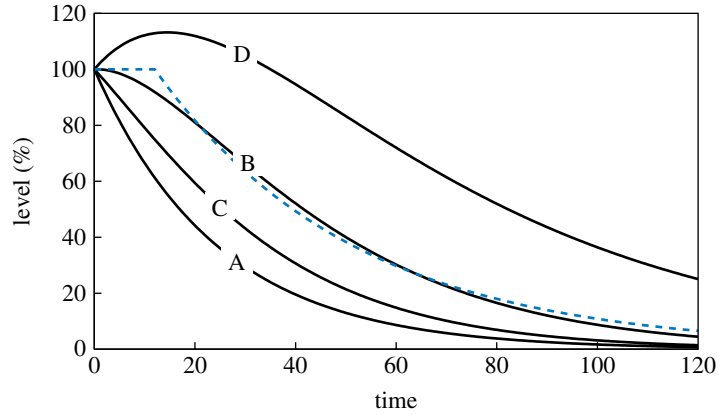

Figure 2.4 *Effect of combination of process lags*

response include steam drum levels, described in Chapter 4, and some schemes for controlling pressure and level in distillation columns, as described in Chapter 12.

We can develop further, for our fired heater example, the concept that order can be thought of as the number of sources of first order lags. For example, if the process operator changes the required valve position, the valve will not move instantaneously to the new position but will approach it with a trajectory close to a first order lag. The structure of the heater has capacitance to absorb heat and so there will be lag, approximating to first order, which will govern how quickly the temperature of the heater coil will increase. Similarly the bulk of fluid inside the coil will cause a lag, as will the thermowell containing the instrument measuring temperature. One could therefore think of the process having an order of four, as illustrated by Figure 2.5. In practice the dynamic behaviour is a product of far more complex transfer of mass and energy. There is no practical way of precisely determining order. What we observe is a *lumped parameter* process most frequently described by deadtime and a single lag. We shall see later, particularly if there is

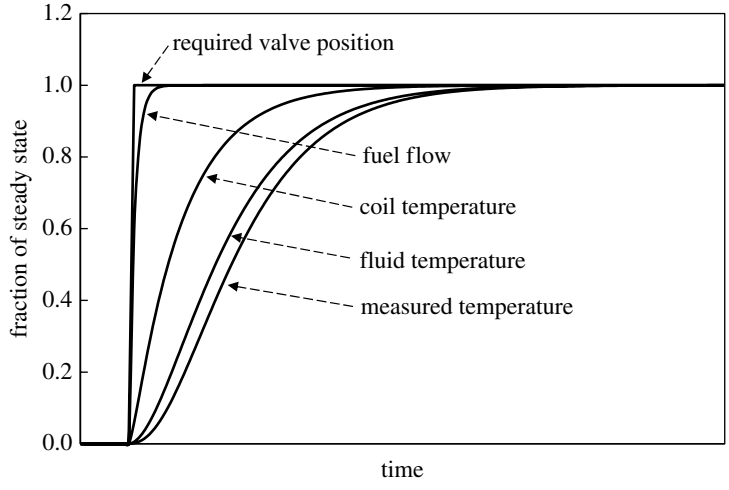

Figure 2.5 *Order of fired heater outlet temperature*

inverse response, that more than one lag might be necessary to describe the behaviour. Indeed, it might be the case that a non-integer number of lags are found to best model the process. However, such *fractional order* models have little practical value in subsequent controller design.

Figures 2.6 to 2.9 show the effect of changing each of these dynamic parameters. Each response is to the same change in MV. Changing K_p has no effect on the behaviour of the process over time. The time taken to reach steady state is unaffected; only the actual steady state changes. Changing θ, τ or n has no effect on actual steady state; only the time taken to reach it is affected. The similarity of the family of curves in Figures 2.8 and 2.9 again shows the principle behind our approximation of first order behaviour – increasing θ has an effect very similar to that of increasing n.

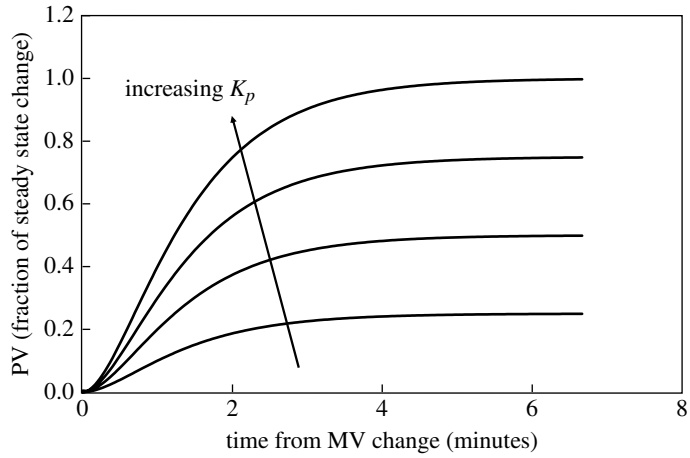

Figure 2.6 *Effect of K_p*

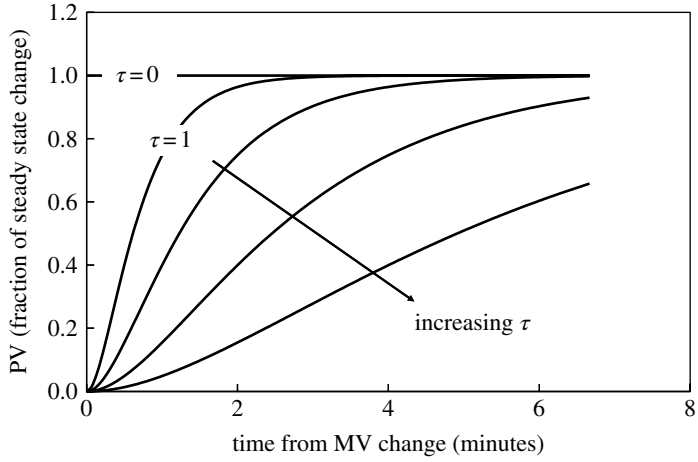

Figure 2.7 *Effect of τ*

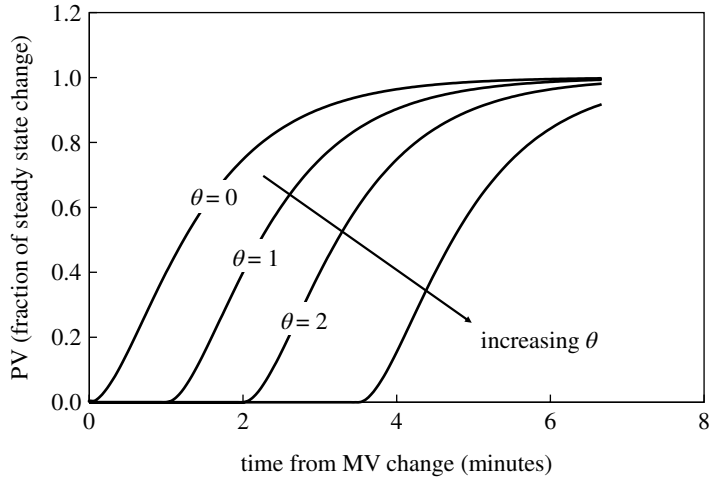

Figure 2.8 *Effect of θ*

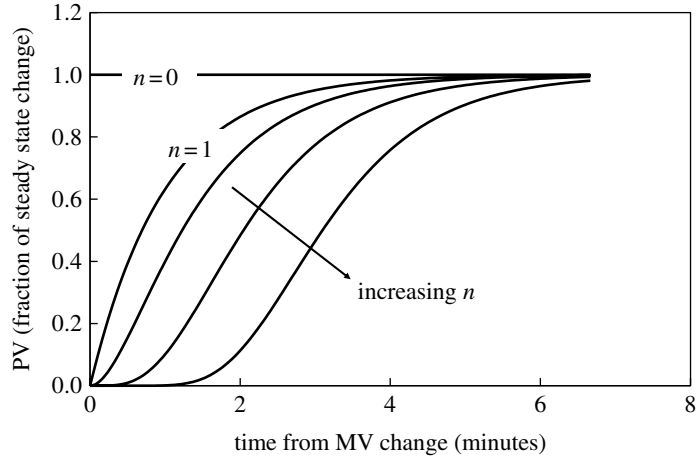

Figure 2.9 *Effect of n (by adding additional lags equal to τ)*

2.2 Cascade Control

Before attempting to determine the process dynamics, we must first explore how they might be affected by the presence of other controllers. One such situation is the use of *cascade control*, where one controller (the *primary* or *master*) adjusts the SP of another (the *secondary* or *slave*). The technique is applied where the process dynamics are such that the secondary controller can detect and compensate for a disturbance much faster than the primary. Consider the two schemes shown in Figure 2.10. If there is a disturbance to the pressure of the fuel header, e.g. because of an increase in consumption on another process, the flow controller will respond quickly and maintain the flow close to SP. As a result, the disturbance to the temperature will be negligible. Without the flow controller, correction will be left to the temperature controller.

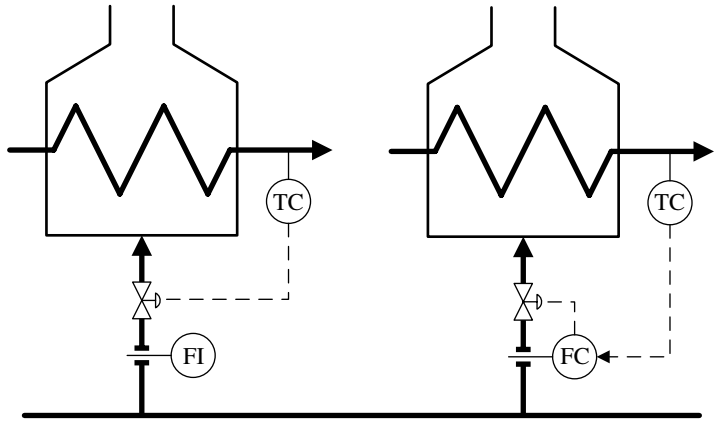

Figure 2.10 *Direct versus cascade control*

But, because of the process dynamics, the temperature will not change as quickly as the flow and nor can it correct as quickly once it has detected the disturbance. As a result, the temperature will deviate from SP for some significant time.

Cascade control also removes any control valve issues from the primary controller. If the valve characteristic is nonlinear, the positioner poorly calibrated or subject to minor mechanical problems, all will be dealt with by the secondary controller. This helps considerably when tuning the primary controller.

Cascade control should not normally be employed if the secondary cannot act more quickly than the primary. In particular, the deadtime in the secondary should be significantly less than that of the primary. Imagine there is a problem with the flow meter in that it does not detect the change in flow for some time. If, during this period, the temperature controller has dealt with the upset then the flow controller will make an unnecessary correction when its measurement does change. This can make the scheme unstable.

Tuning controllers in cascade should always be completed from the bottom up. Firstly, the secondary controller will on occasions be in use without the primary. There may, for example, be a problem with the primary or its measurement may be out of range during start-up or shutdown of the process. We want the secondary to perform as effectively as possible and so it should be optimally tuned as a standalone controller. The second reason is that the MV of the primary controller is the SP of the secondary. When performing step tests to tune the primary we will make changes to this SP. The secondary controller is now effectively part of the process and its tuning will affect the dynamic relationship between the primary PV and MV. If, after tuning the primary, we were to change the tuning in the secondary then the tuning in the primary would no longer be optimum.

Cascade control, however, is not the only case where the sequence of controller tuning is important. In general, before performing a plant test, the engineer should identify any controllers that will take corrective action during the test itself. Any such controller should be tuned first. In the case of cascade control, clearly the secondary controller takes corrective action when its SP is changed. But consider the example shown in Figure 2.11. The heater has a simple flue gas oxygen control which adjusts a damper to maintain the required excess air. When the downward step is made to the fuel flow SP the oxygen controller, if in automatic mode, will take corrective action to reduce the air rate and return the oxygen content to SP. However, if this controller is in manual mode, no corrective action is taken, the oxygen level will rise and the heater efficiency will fall. As a result the heater outlet temperature will fall by more than it did in the first test. Clearly this affects the process gain between temperature and fuel. Imagine now that the oxygen control

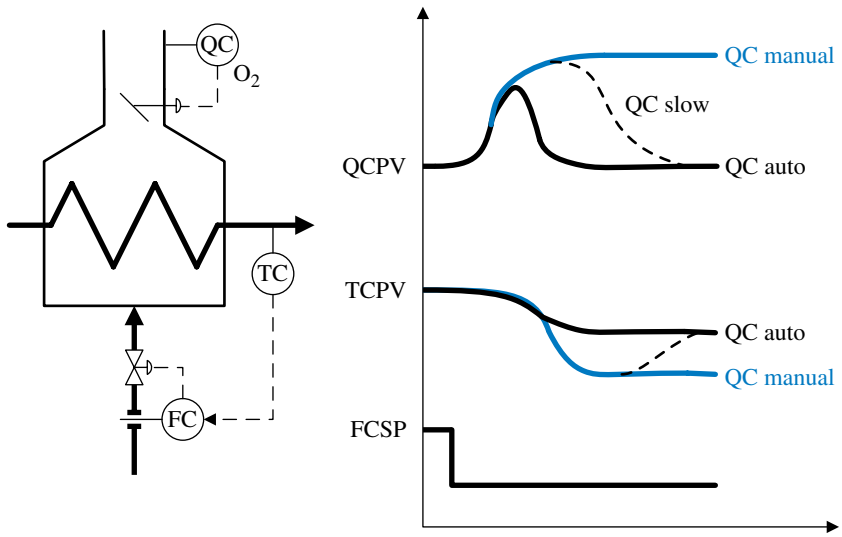

Figure 2.11 *Effect of other controllers*

is re-tuned to act more slowly. The dynamic behaviour of the temperature with respect to fuel changes will be quite different. So we have the situation where an apparently unrelated controller takes corrective action during the step test. It is important therefore that this controller is properly tuned before conducting the test.

In the case of testing to support the design of MPC, the MVs are likely to be mainly basic controllers and it is clear that these controllers should be well-tuned before starting the step tests. However, imagine that one of the MVs is the feed flow controller. When its SP is stepped there is likely to be a large number of regulatory controllers that will take corrective action during the test. Many of these will not be MVs but nevertheless need to be properly tuned before testing begins.

2.3 Model Identification

Model identification is the process of quantifying process dynamics. The techniques available fall into one of two approaches – *open loop* and *closed loop* testing. Open loop tests are performed with either no controller in place or, if existing, with the controller in *manual mode*. A disturbance is injected into the process by directly changing the MV. Closed loop tests are conducted with the controller in *automatic mode* and may be used when an existing controller provides some level (albeit poor) of stable control. Under these circumstances the MV is changed indirectly by making a change to the SP of the controller. When first introduced to closed loop testing, control engineers might be concerned that the tuning of the controller will affect the result. While a slower controller will change the MV more slowly, it does not affect the relationship between PV and MV. It is this relationship we require for model identification – not the relationship between PV and SP.

Such plant testing should be well organised. While it is clear that the process operator must agree to the test, there needs to be discussion about the size and duration of the steps. It is in the engineer's interest to make these as large as possible. The operator of course would prefer no disturbance be made. The operator also needs to appreciate that other changes to the process should not be made during the test. While it is

possible to determine the dynamics of simultaneous changes to several variables, the analysis is complex and more prone to error.

It seems too obvious to state that the process instrumentation should be fully operational.

Many data historians included a compression algorithm to reduce the storage requirement. When later used to recover the original data, some distortion will occur. While this is not noticeable in most applications, such as process performance monitoring and accounting, it can affect the apparent process dynamics. Any compression should therefore be disabled prior to the plant tests. Indeed, it is becoming increasingly common to disable compression for all data. The technique was developed to reduce data storage costs but, with these falling rapidly, has become unnecessary. Removing it completely means that historical data collected during a routine SP change can be used for model identification, obviating the need for a step test.

It is advisable to collect more than just the PV and MV. If the testing is to be done closed loop then the SP should also be recorded. Any other process parameter which can cause changes in the PV should also be collected. This is primarily to ensure that they have not changed during the testing, or to help diagnose a poor model fit. While such disturbances usually invalidate the test, it may be possible to account for them and so still identify an accurate model.

Ideally, testing should be planned for when there are no other scheduled disturbances. It can be a good idea to avoid shift changeovers – partly to avoid having to persuade another crew to accept the process disturbances but also to avoid the changes to process conditions that operators often make when returning from lengthy absences. If ambient conditions can affect the process then it is helpful to avoid testing when these are changing rapidly, e.g. at dawn or dusk and during rainstorms. Testing should also be scheduled to avoid any foreseen changes in feed composition or operating mode.

Laboratory samples are often collected during plant tests. These are usually to support the development of *inferential properties* (as described in Chapter 9). Indeed steady operation, under conditions away from normal operation, can provide valuable data 'scatter'. Occasionally series of samples are collected to obtain dynamic behaviour, for example, if an on-stream analyser is temporarily out of service or its installation delayed. The additional laboratory testing generated may be substantial compared to the normal workload. If the laboratory is not expecting this, then analysis may be delayed for several days with the risk that the samples may degrade.

The most accurate way of determining the dynamic constants is by a computer-based curve fitting technique which uses the values of the MV and PV collected frequently throughout the test. If we assume that the process can be modelled as first order plus deadtime (*FOPDT*), then in principle this involves fitting Equation (2.16) to the data collected.

$$PV_n = a_0 + a_1 PV_{n-1} + b_1 MV_{n-\theta/ts} \tag{2.16}$$

$$a_0 = \left(1 - e^{-ts/\tau}\right) bias \qquad a_1 = e^{-ts/\tau} \qquad b_1 = K_p\left(1 - e^{-ts/\tau}\right) \tag{2.17}$$

PV_n and PV_{n-1} are the current and previous predicted values with PV_0 set to the actual starting PV. The *bias* term is necessary because *PV* will not generally be zero when *MV* is zero. Care must be taken to ensure that the data collection interval (ts) has the same units of time as the process lag (τ) and deadtime (θ).

The values of K_p, θ, τ and *bias* are adjusted to minimise the sum of the squares of the error between the predicted PV and the actual PV. When θ is not an exact multiple of the data collection interval (ts), then the MV is interpolated between the two values either side of the required value.

$$MV_{n-\theta/ts} = MV_{n-\mathrm{int}(\theta/ts)} - \left(\frac{\theta}{ts} - \mathrm{int}\left(\frac{\theta}{ts}\right)\right)\left(MV_{n-\mathrm{int}(\theta/ts)} - MV_{n-\mathrm{int}(\theta/ts)-1}\right) \tag{2.18}$$

An alternative approach is to first choose a value for θ that is an integer multiple of ts and apply linear regression to identify a_0, a_1, b_1 and b_2 in the equation

$$PV_n = a_0 + a_1 PV_{n-1} + b_1 MV_{n-\theta/ts} + b_2 MV_{n-\theta/ts-1} \tag{2.19}$$

An iterative approach is then followed to find the best integer value of θ/ts. Once the coefficients are known, then K_p can be derived:

$$K_p = \frac{b_1 + b_2}{1 - a_1} \tag{2.20}$$

The deadtime is determined from the best integer value of θ/ts by

$$\theta = \left[\text{int}\left(\frac{\theta}{ts} \right) + \frac{b_2}{b_1 + b_2} \right] ts \tag{2.21}$$

Process lag is derived by rearranging Equation (2.17).

$$\tau = \frac{-ts}{\ln(a_1)} \tag{2.22}$$

There are two other approaches described in some texts. The first begins with describing the process as a differential equation, similar to Equation (2.6).

$$\tau \frac{dPV}{dt} + PV = K_p MV_{t-\theta} \tag{2.23}$$

Writing this as its discrete approximation gives

$$\tau \left(\frac{PV_n - PV_{n-1}}{ts} \right) + PV_{n-1} = K_p MV_{n-\theta/ts} \tag{2.24}$$

Rearranging

$$PV_n = \left(1 - \frac{ts}{\tau} \right) PV_{n-1} + K_p \frac{ts}{\tau} MV_{n-\theta/ts} \tag{2.25}$$

Comparison to Equation (2.16) gives

$$a_1 = 1 - \frac{ts}{\tau} \qquad \text{and} \qquad b_1 = K_p \frac{ts}{\tau} \tag{2.26}$$

The same result could be achieved by making the *first order Taylor approximation* in Equation (2.17).

$$e^{-ts/\tau} = 1 - \frac{ts}{\tau} \tag{2.27}$$

This approximation slightly reduces the accuracy of the result. The other approach often published defines the coefficients as

$$a_1 = \frac{1}{1+\left(\frac{ts}{\tau}\right)} \qquad \text{and} \qquad b_1 = K_p \frac{\left(\frac{ts}{\tau}\right)}{1+\left(\frac{ts}{\tau}\right)} \tag{2.28}$$

These are developed by applying the first order Taylor approximation to the reciprocal of the exponential function.

$$e^{-ts/\tau} = \frac{1}{e^{ts/\tau}} = \frac{1}{1+\left(\frac{ts}{\tau}\right)} \tag{2.29}$$

While a slightly more accurate approximation than Equation (2.26), it remains an unnecessary approximation.

More complex equations can be used to identify higher order models. As we show in Chapter 14, the following equation permits the identification of a second order model with (if present) inverse response or overshoot.

$$PV_n = a_0 + a_1 PV_{n-1} + a_2 PV_{n-2} + b_1 MV_{n-\theta/ts} + b_2 MV_{n-\theta/ts-1} + b_3 MV_{n-\theta/ts-2} \tag{2.30}$$

It may not be possible to convert this to *parametric form* (using K_p, θ, τ_1, τ_2 and τ_3) and so it is likely to be more convenient to use the model as identified.

This model identification technique can be applied to both open and closed loop tests. Multiple disturbances are made in order to check the repeatability of the results and to check linearity. While not necessary for every step made, model identification will be more reliable if the test is started with the process as steady as possible and allowed to reach steady state after at least some of the steps.

The data collection interval can be quite large. We will show later that steady state is virtually reached within $\theta + 5\tau$. Assuming we need around 30 points to achieve a reasonably accurate fit and that we make both an increase and a decrease in the MV, then collecting data at a one-minute interval would be adequate for a process which has time constants of around 2 or 3 minutes. If more than two steps are performed, as would usually be the case, dynamics less than a minute can be accurately identified.

It is important to avoid *correlated steps*. Consider the series of steps shown in Figure 2.12. There is clearly a strong correlation between the PV and the MV, with K_p of 1.0 and θ of around 3.0 minutes. However, there is theoretically an equally accurate model with K_p of –1.0 and θ of around 33.0 minutes. Real process data will contain some random unaccounted variation and so there is a probability of 50% that the wrong model appears to be the more accurate. Performing a series of steps of varying size and duration, as in Figure 2.13, would avoid this problem. Indeed, this is the principle of the *pseudo-random binary sequence (PRBS)* used by some automatic step-testing products. This comprises a series of steps, of user-specified amplitude, made at an apparently random interval in alternating directions.

If testing is performed with an existing controller in automatic mode it is similarly important that it exhibits no oscillatory behaviour. Even an oscillation decaying in amplitude can cause problems similar to those that arise from correlated steps. Even if testing is conducted with the controller in manual mode, oscillatory behaviour caused by an apparently unrelated controller can also give problems. Since the

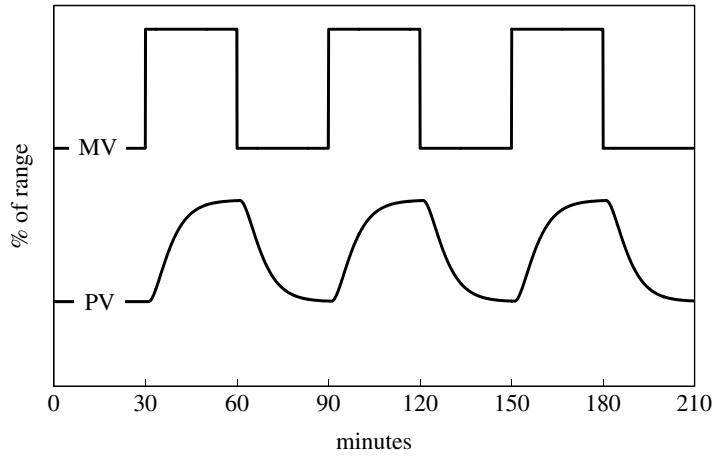

Figure 2.12 *Correlated steps*

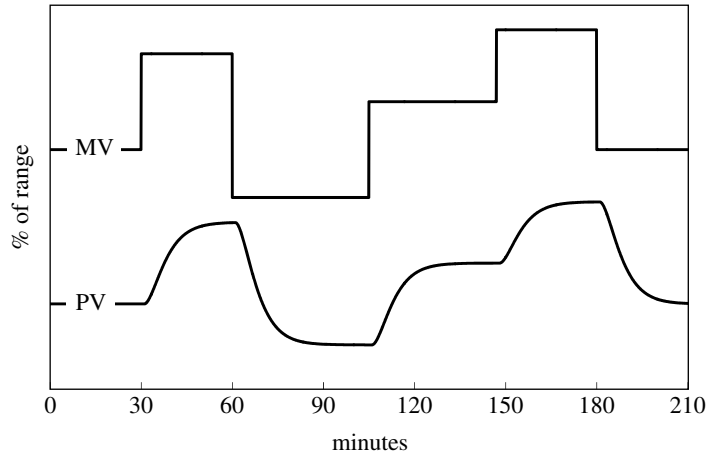

Figure 2.13 *Non-correlated steps*

purpose of step-testing is to re-tune the controller, there is nothing to be lost by making a large reduction to controller gain, sufficient to make it stable, before conducting the test.

Model identification software packages will generally report some measure of confidence in the model identified. A low value may have several causes. Firstly, noise in either the MV or PV, if of a similar order of magnitude to the changes made, can disguise the model.

Secondly, if the MV is a valve or similar actuator, problems such as *stiction* and *hysteresis* will reduce model accuracy. These are shown in Figure 2.14. Stiction (or *static friction*), caused by excessive friction, requires that the signal change to start the valve moving is greater than the signal to keep it moving. Thus a small change in the signal may have no effect on the PV, whereas a subsequent change will affect it as expected. This is also known as *stick-slip*. Figure 2.15 shows a typical symptom in a controller. Oscillation, also known as *hunting*, occurs when only small changes to valve position are required, i.e. when the PV is

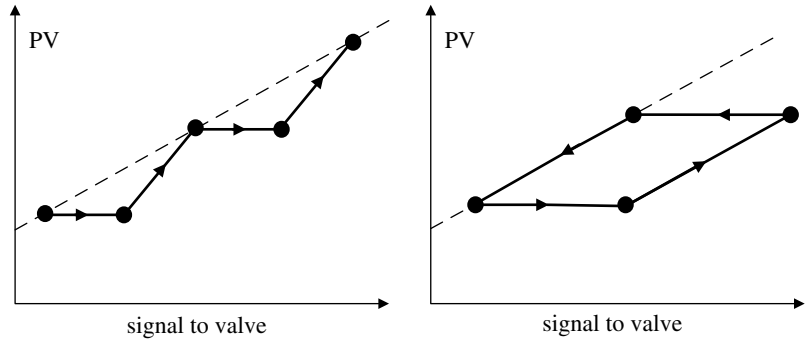

Figure 2.14 *Stiction and hysteresis in a control valve*

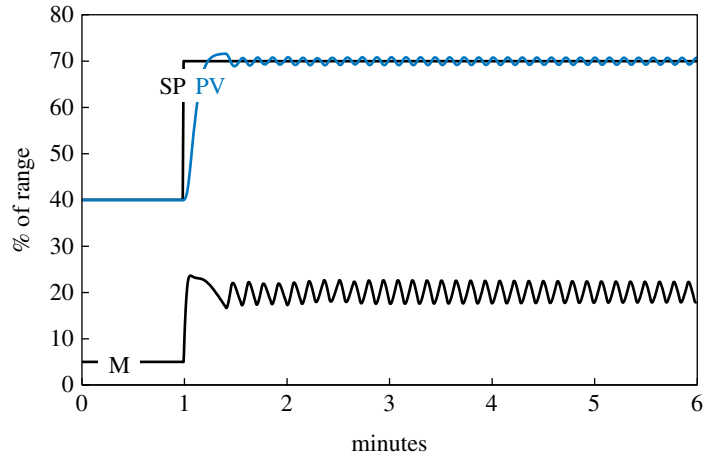

Figure 2.15 *Hunting caused by stiction in closed loop*

close to the SP. Once stiction is overcome by the controller's integral action, the resulting change in valve position is greater than required and the PV overshoots the SP. Since there are other potential causes for the oscillation, stiction is better diagnosed by an open loop test. This comprises making a series of small increases to the signal to the valve, followed by a series of small decreases. The result of this is shown in Figure 2.16. The first one or two increases have no effect until the total increase overcomes the stiction. A similar pattern is followed when decreasing the signal. Clearly, if such tests were being made to obtain process dynamics, it is unlikely that a true estimate of process gain would be possible.

As described in Chapter 3, oscillatory behaviour can also be caused by the controller being too tightly tuned. Misdiagnosing the problem would result in attempting to solve it by adjusting the tuning to make the controller slower. Following a disturbance, the controller will take then longer to overcome stiction – resulting in a larger deviation from SP before the control valve moves. This will result in a reduction in the frequency of oscillation and an increase in its amplitude. Reducing the controller gain, by a factor of 4, caused the performance to change from that shown by Figure 2.15 to that in Figure 2.17.

Hysteresis (sometimes called *backlash* or *deadband*) is usually caused by wear in couplings and bearings, resulting in some clearance between contacting parts and creating play in the mechanism. As the

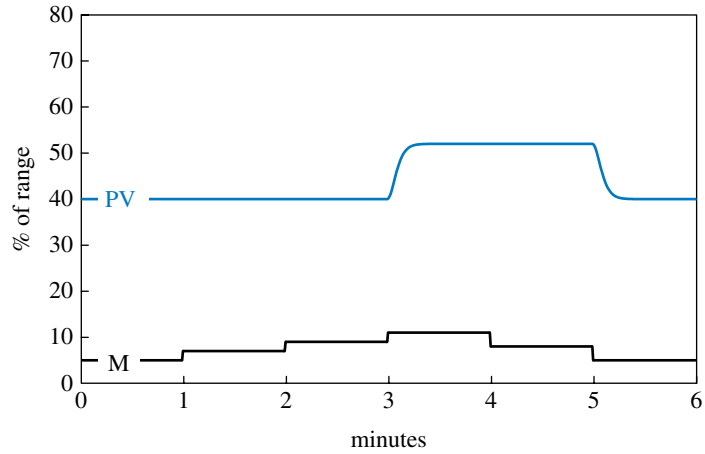

Figure 2.16 *Effect of stiction in an open loop test*

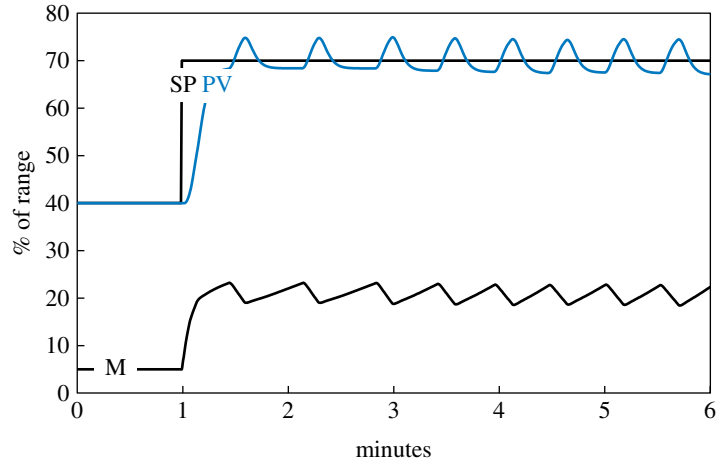

Figure 2.17 *Effect on stiction of reducing controller gain*

signal is increased, this play is first overcome before the actuator begins to move. It will then behave normally until the signal is reversed, when the play must again be overcome. Figure 2.18 shows an example of a real case of valve hysteresis where the level controller performed very badly. The level PV and controller output are trended over 6 hours in Figure 2.19. The behaviour is explained by Figure 2.20 which shows the relationship between flow and level controller output. The coloured region comprises 5,000 values collected from the plant historian at a one-minute interval. The black line shows the route typically taken by the controller during the approximate 45 minute period of oscillation. In this severe case hysteresis, following a reversal of direction, required the controller output to move by about 35% before the valve began to move.

Figure 2.21 shows the erratic behaviour, caused by hysteresis, of another controller output during SP changes. Under these circumstances the problem may not be immediately obvious – particularly as the PV

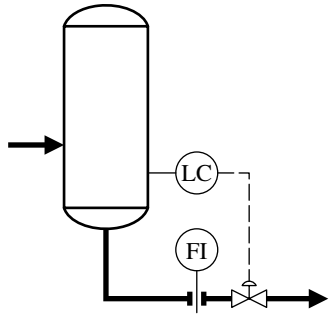

Figure 2.18 Example of process showing hysteresis

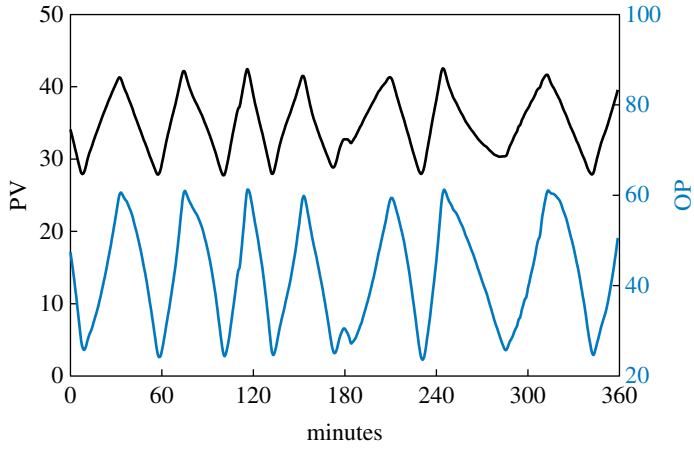

Figure 2.19 Performance of level controller

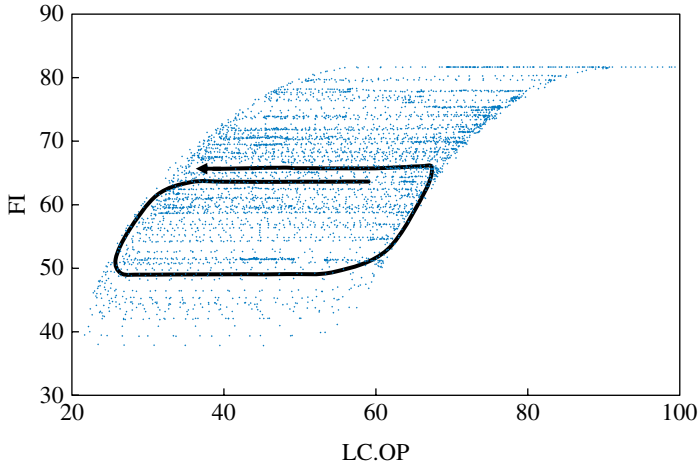

Figure 2.20 Relationship between flow and controller output

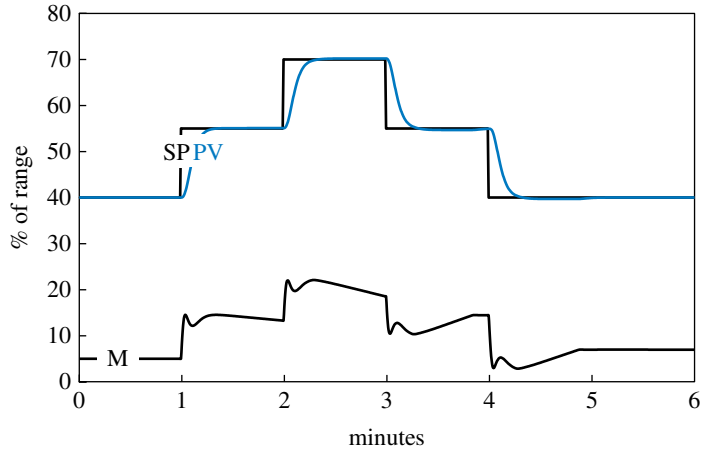

Figure 2.21 *Erratic behaviour caused by hysteresis during SP changes*

seems to be well controlled. However, closer inspection shows that, although the SP is returned to its starting value, the controller output (*M*) does not. Again it is unlikely that a reliable dynamic model could be identified from the closed loop test. Figure 2.22 shows an open loop test, conducted in the same way as that to identify stiction. Each of the steps in *M* is of the same size but the steady state changes in *PV* clearly are not. The lack of consistency would make impossible an accurate estimate of process gain. The dashed line, showing the true valve position, explains the behaviour.

Thirdly, the relationship between PV and MV may be inherently nonlinear. Some model identification packages can analyse this. If not, then plotting historically collected steady-state values of PV against MV will permit linearity to be checked and possibly a linearising function developed. Techniques for this are covered in Chapter 5.

While computer-based packages are readily available, a great deal of attention is given in text books to manual analysis of step-tests. Much of the remainder of this section describes those commonly

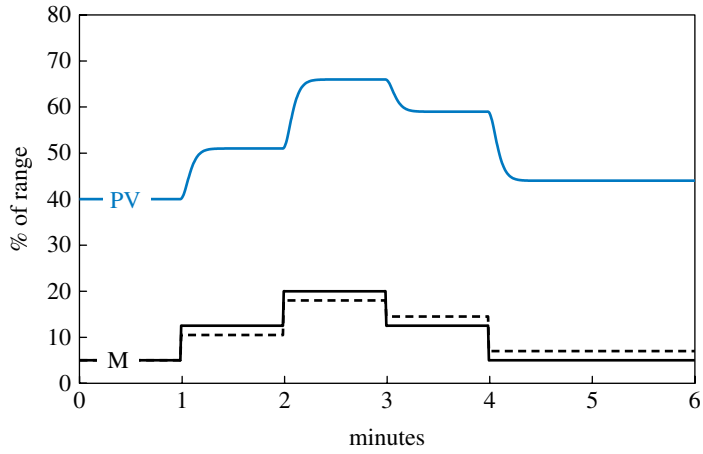

Figure 2.22 *Effect of hysteresis in open loop test*

published – primarily to draw attention to their limitations. With most processes now having some form of process data historisation, it is unlikely that the engineer should ever see the need to apply them – except perhaps to make a quick visual estimate of the process dynamics. The techniques can also only be used to identify first order plus deadtime models and the MV must be changed as a single step, starting and ending at steady state. This is not always possible for any of several reasons.

- Any existing controller will need to be switched to manual mode. This may be undesirable on an inherently unstable process.
- There are many processes which rarely reach true steady state and so it would be optimistic to start and finish the test under this condition.
- The size of the step must be large enough to have a noticeable effect on the process. If the PV is subject to noise, small disturbances will be difficult to analyse accurately. The change in PV needs to be at least five times larger than the noise amplitude. This may cause an unacceptable process disturbance and instead several smaller steps (or perhaps a ramp change) may be necessary.
- Dynamics, as we shall see later in Chapter 6, are not only required for changes in the MV but also for *disturbance variables (DV)*. It may be that these cannot be changed as steps. For example, ambient temperature, if to be included as a DV, clearly cannot be stepped.

If a single step is practical it will still be necessary to conduct multiple tests, analysing each separately, to confirm repeatability and to check for linearity.

The most widely published method is based on the principle that a process with zero deadtime will complete 63.2% of the steady state change within one process lag. If, in Equation (2.9), we set t equal to τ, we get

$$x = 0.632 x_{new} \tag{2.31}$$

This calculation can be repeated for multiples of τ, resulting in the graph shown in Figure 2.23. While, in theory, the process will never truly reach steady state, within five time constants it will be very close - having completed 99.3% of the change.

In general, however, we have to accommodate deadtime in our calculation of dynamics. Ziegler and Nichols [1] proposed the method using the *tangent of steepest slope*. Shown in Figure 2.24, it involves

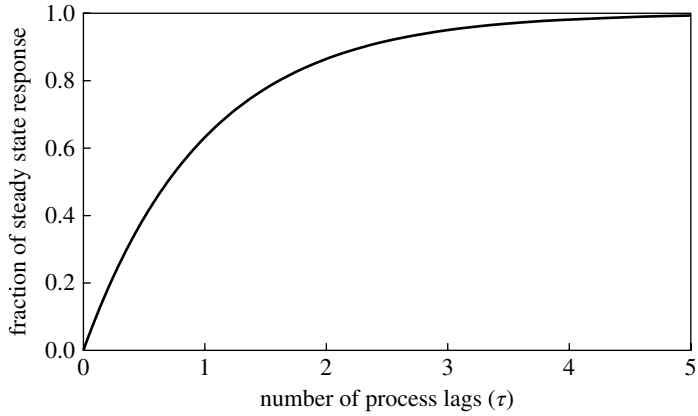

Figure 2.23 *Time to reach steady state*

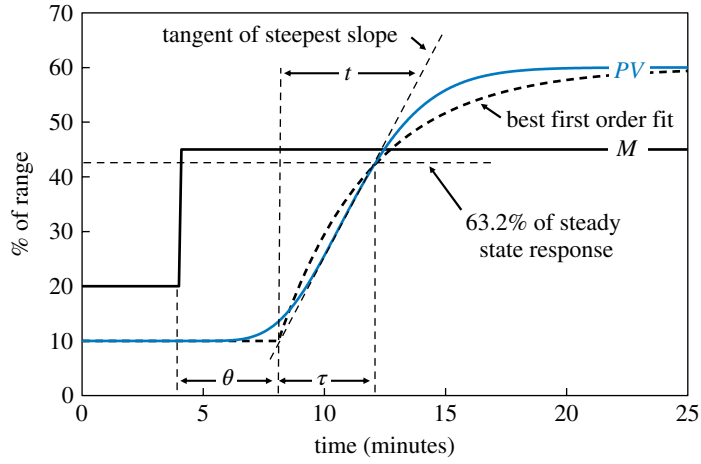

Figure 2.24 *Ziegler–Nichols steepest slope method*

identifying the point at which the PV is changing most rapidly and then drawing a tangent to the curve at this point. Where it crosses the value of the PV at the start of the test gives the process deadtime (θ). There are two methods for determining the process lag (τ). While not mentioned by Ziegler and Nichols, the time taken to reach 63.2% of the steady state response is $\theta + \tau$, so once θ is known, τ can be derived. Ziegler and Nichols, as we shall see later when looking at their controller tuning method, instead characterised the process by determining the slope of the tangent (R). We will show later that this is equivalent to defining τ as the distance labelled t in Figure 2.24. For a truly first order process with deadtime this will give the same result. For higher order systems this approach is inaccurate. K_p is determined from Equation (2.2).

The resulting first order approximation is included in Figure 2.24. The method forces it to pass through three points – the intersection of the tangent with the starting PV, the 63.2% response point and the steady state PV. In this example θ is estimated at 4.2 minutes and τ as 3.8 minutes. The method is practical but may be prone to error. Correctly placing the line of steepest slope may be difficult – particularly if there is measurement noise. Drawing it too steeply will result in an overestimate of θ and an underestimate of τ. The ratio θ/τ (in this case 1.11), used by most controller tuning methods, would thus much larger than the true value.

An alternative approach is to identify two points on the response curve. A first order response is then forced through these two points and the steady-state values of the PV. Defining t_a as the time taken to reach a% of the steady-state response and t_b as the time taken to reach b%, the process dynamics can be derived from the formulae

$$\tau = \frac{t_b - t_a}{\ln\left(1 - \frac{a}{100}\right) - \ln\left(1 - \frac{b}{100}\right)} \tag{2.32}$$

$$\theta = t_a + \tau.\ln\left(1 - \frac{a}{100}\right) \tag{2.33}$$

By manipulating these equations, θ may also be derived from either of the following:

$$\theta = t_b + \tau . \ln\left(1 - \frac{b}{100}\right) \tag{2.34}$$

$$\theta = \frac{t_a \ln\left(1 - \frac{b}{100}\right) - t_b . \ln\left(1 - \frac{a}{100}\right)}{\ln\left(1 - \frac{b}{100}\right) - \ln\left(1 - \frac{a}{100}\right)} \tag{2.35}$$

The values of a and b need not be symmetrical but, for maximum accuracy, they should not be close together nor too close to the start and finish steady-state conditions. Choosing values of 25% and 75% reduces Equations (2.32) and (2.33) to

$$\tau = 0.91\left(t_{75} - t_{25}\right) \tag{2.36}$$

$$\theta = t_{25} - 0.29\tau \tag{2.37}$$

Figure 2.25 shows the application of this method and the resulting first order approximation. In this case θ is estimated at 4.7 minutes, τ as 3.0 minutes and θ/τ as 1.59. Others have used different points on the curve. For example, Sundaresan and Krishnaswamy [2] chose

$$\tau = 0.68\left(t_{85.3} - t_{35.3}\right) \tag{2.38}$$

$$\theta = 1.29t_{35.3} - 0.29t_{85.3} \tag{2.39}$$

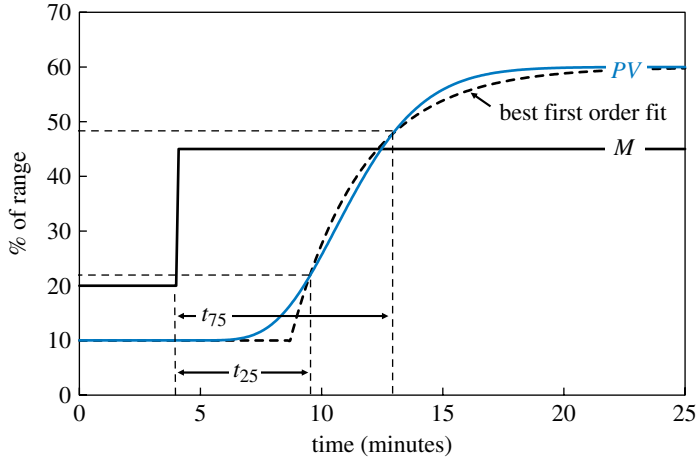

Figure 2.25 *Two-point method*

Applying this method gives θ as 5.2 minutes, τ as 2.5 minutes and θ/τ as 2.06. Smith [3] proposed

$$\tau = 1.50\left(t_{63.2} - t_{28.3}\right) \tag{2.40}$$

$$\theta = t_{63.2} - \tau \tag{2.41}$$

Applying this method gives θ as 4.7 minutes, τ as 3.2 minutes and θ/τ as 1.47.

Another approach is to use more points from the curve and apply a least squares technique to the estimation of θ and τ. Rearranging Equation (2.33) we get

$$t_a = \theta + \tau \ln\left(\frac{100}{100 - a}\right) \tag{2.42}$$

So, by choosing points at 10% intervals

$$t_{10} = \theta + 0.1054\tau \tag{2.43}$$

$$t_{20} = \theta + 0.2231\tau \tag{2.44}$$

$$t_{30} = \theta + 0.3567\tau \tag{2.45}$$

$$t_{40} = \theta + 0.5108\tau \tag{2.46}$$

$$t_{50} = \theta + 0.6931\tau \tag{2.47}$$

$$t_{60} = \theta + 0.9163\tau \tag{2.48}$$

$$t_{70} = \theta + 1.2040\tau \tag{2.49}$$

$$t_{80} = \theta + 1.6094\tau \tag{2.50}$$

$$t_{90} = \theta + 2.3026\tau \tag{2.51}$$

Using a spreadsheet package, θ and τ would be adjusted to minimise the sum of the square of the errors between the actual time to reach each % of steady-state and the time predicted by each of the Equations (2.43) to (2.51). Applying this method gives θ as 5.0 minutes, τ as 2.7 minutes and θ/τ as 1.83.

The PV curve shown in Figures 2.24 and 2.25 actually comprises 250 points collected at 6-second intervals. Least squares regression using all of these values results in a 'best' estimate for θ and τ of 4.7 and 2.9 minutes respectively with θ/τ as 1.62. This is illustrated in Figure 2.26. The curves cross at about 25% and 77% of the steady state response showing, that if a two-point method is to be used, these would be the best choice (in this example). While this will not be exactly the case for all processes it is reasonable to assume that using 25% and 75% would, in general, be a reliable choice. Remembering that process dynamics are unlikely to remain constant, the advantage of ensuring precision with a single test is perhaps small. However, the error of around 30% in estimating θ/τ, arising from applying either the Ziegler–Nichols steepest slope method or the two-point method proposed by Sundaresan and Krishnaswamy, is excessive.

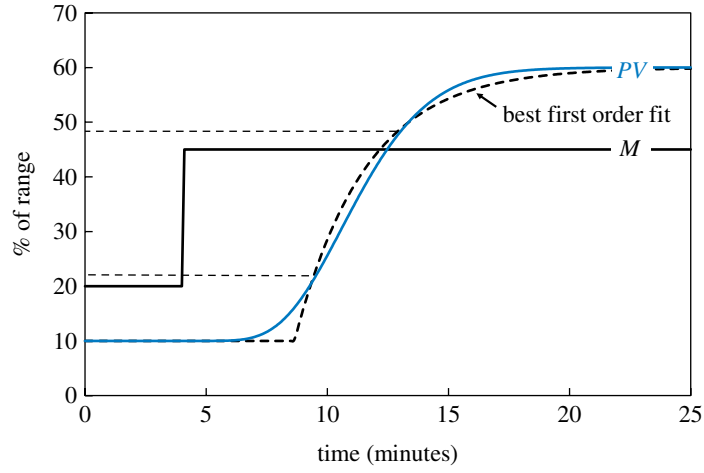

Figure 2.26 *'Best' least squares fit*

With any model identification technique care should be taken with units. As described earlier in this chapter, K_p should be dimensionless if the value is to be used in tuning a DCS-based controller. The measurements of PV and MV, used in any of the model identification techniques described, should first be converted to fractions (or %) of instrument range. For computer-based MPC, K_p would usually be required in engineering units; so no conversion should be made. Both θ and τ should be in units consistent with the tuning constants. It is common for the integral time (T_i) and the derivative time (T_d) to be in minutes, in which case the process dynamics should be in minutes; but there are systems which use seconds and so the dynamics should then be determined in seconds.

Figure 2.27 shows the effect of increasing order, but unlike Figure 2.9, by adjusting the time constants so that the overall lag remains the same, i.e. all the responses reach 63% of the steady state change after one minute. It shows that, for large values of n, the response becomes closer to a step change. This confirms that

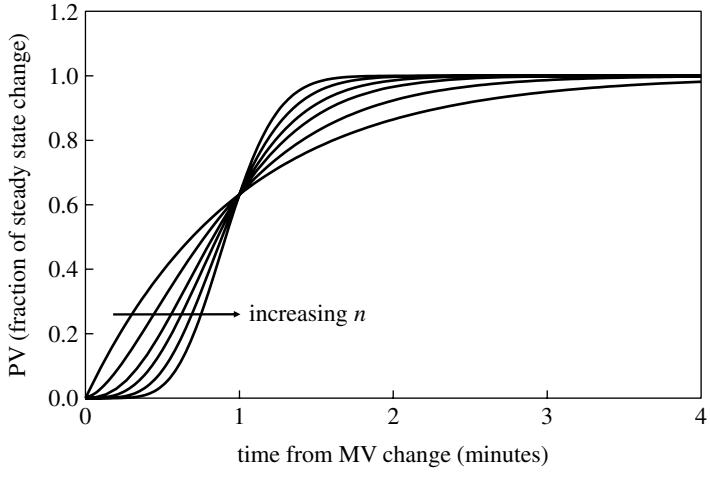

Figure 2.27 *Effect of n (by keeping 63% response time equal)*

a series of lags can be approximated by deadtime. But it also means that deadtime can be approximated by a large number of small lags. We will cover, in Chapters 6, 7 and 8, control schemes that require a deadtime algorithm. If this is not available in the DCS, then this approximation would be useful.

2.4 Integrating Processes

The fired heater that we have worked with is an example of a *self-regulating* process. Following the disturbance to the fuel valve the temperature will reach a new steady state without any manual intervention. For example, an increase in valve opening will cause the temperature to rise until the additional energy leaving in the product is equal to the additional energy provided by the fuel. Not all processes behave this way. For example, if we were trying to obtain the dynamics for a future level controller, we would make a step change to the manipulated flow. If this is the flow leaving a vessel and the inlet flow is kept fixed, the level would not reach a new steady state unless some intervention is made. This *non-self-regulating* process can also be described as an *integrating* process.

While level is the most common example there are many others. For example, many pressure controllers show a similar behaviour. Pressure is a measure of the inventory of gas in a system, much like a level is a measure of liquid inventory. An imbalance between the gas flow into and out of the system will cause the pressure to ramp without reaching a new steady state. However, not all pressures show pure integrating behaviour. For example, if the flow in or out of the system is manipulated purely by valve position, i.e. no flow control, then the resulting change in pressure will cause the flow through the valve to change until a new equilibrium is reached. Even with flow controllers in place, if flow is measured by an uncompensated orifice-type meter, the error created in the flow measurement by the change in pressure will also cause the process to be self-regulating.

Some temperatures can show integrating behaviour. If increasing heater outlet temperature also causes the heater inlet temperature to rise, through some recycle or heat integration, then the increase in energy input will cause the outlet temperature to ramp up.

The response of a typical integrating process is shown as Figure 2.28. Since it does not reach steady state, we cannot immediately apply the same method of determining the process gain from the steady-state change

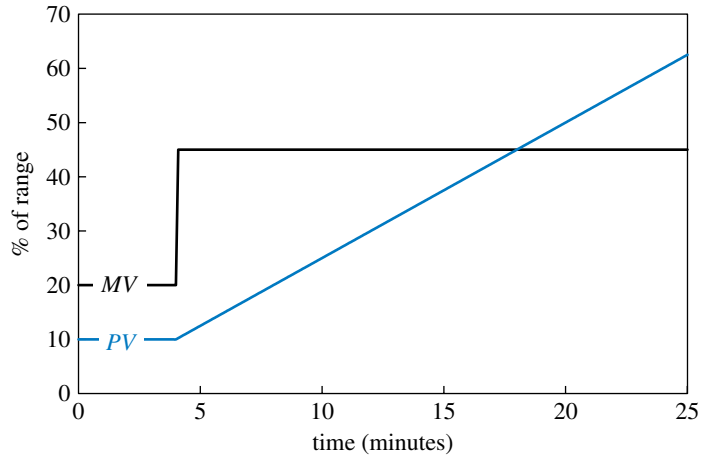

Figure 2.28 *Integrating process*

in PV. Nor can we use any technique which relies on a percentage approach to steady state. By including a bias (because it is not true that the PV is zero when the MV is zero), we can modify Equation (2.2) for a self-regulating process to

$$PV = K_p.MV + bias \tag{2.52}$$

In the case of an integrating process, the PV also varies with time, so we describe it by

$$PV = K_p \int \left(MV + bias\right) dt \tag{2.53}$$

or, by differentiating,

$$\frac{dPV}{dt} = K_p \left(MV + bias\right) \tag{2.54}$$

By replacing *PV* with its derivative we can therefore apply the same model identification techniques used for self-regulating processes.

While, for DCS-based controllers, *PV* and *MV* remain dimensionless K_p must now have the units of reciprocal time. The units will depend on whether rate of change of *PV* is expressed in sec^{-1}, min^{-1} or hr^{-1}. Any of these may be used, provided consistency is maintained. Throughout this book we will use min^{-1}.

We can omit the lag term when characterising the process dynamics of an integrating process. Although the process is just as likely to include a lag, this manifests itself as deadtime. Figure 2.29 illustrates the effect of adding lag to the PV. In this case, a lag of 3 minutes has caused the apparent deadtime to increase by about the same amount. After the initial response the PV trend is still a linear ramp.

We can thus characterise the response using only K_p and θ. These can be derived by fitting Equation (2.16) with the coefficients.

$$a_0 = K_p.ts.bias \qquad a_1 = 1 \qquad b_1 = K_p.ts \tag{2.55}$$

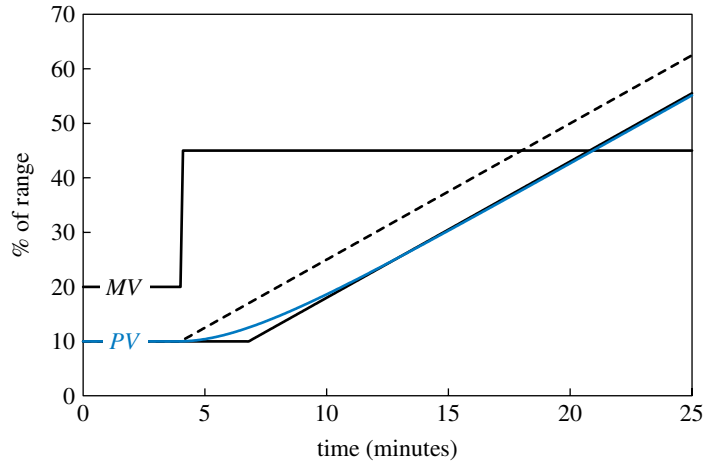

Figure 2.29 *Effect of lag on an integrating process*

If deadtime cannot be assumed to be an integer number of scan intervals then we can fit the equivalent of Equation (2.19), i.e.

$$PV_n = a_0 + PV_{n-1} + b_1 MV_{n-\theta/ts} + b_2 MV_{n-\theta/ts-1} \tag{2.56}$$

Deadtime (θ) is derived by applying Equation (2.21); other parameters from:

$$K_p = \frac{b_1 + b_2}{ts} \tag{2.57}$$

$$bias = \frac{a_0}{b_1 + b_2} \tag{2.58}$$

It is common, when designing MPC, to substitute approximate integrating models for self-regulating processes that have very large lags. By not having to wait for steady state, this reduces the time taken by step-testing. It also simplifies MPC by reducing the number of sample periods required in the controller to predict the future value of the PV. While this approach generally works well with MPC, it must be applied with caution to basic controllers. The principle behind the approximation is to determine the slope of the process response curve at the point where the deadtime has elapsed. The behaviour of a first order self-regulating process can be described by Equation (2.59), where t is the time elapsed since the expiry of the deadtime.

$$PV = K_p\left(1 - e^{-t/\tau}\right)\left(MV + bias\right) \tag{2.59}$$

Differentiating gives

$$\frac{dPV}{dt} = \frac{K_p}{\tau} e^{-t/\tau}\left(MV + bias\right) \tag{2.60}$$

When t is zero

$$\frac{dPV}{dt} = \frac{K_p}{\tau}\left(MV + bias\right) \tag{2.61}$$

Comparison with Equation (2.54) shows that this describes an integrating process with a process gain K'_p, where

$$K'_p = \frac{K_p}{\tau} \tag{2.62}$$

We will show in Chapter 3 that many tuning methods, that give tuning calculations for both self-regulating and integrating processes, do so by applying this approximation. However, for the preferred method, this can fail under certain circumstances. To demonstrate this, the preferred tuning method was used to design controllers for a wide range of self-regulating processes. The process gain (K'_p) of the approximated integrating process was then adjusted to obtain the best possible response to a SP change with the same controller in place. The error between this value and that predicted by Equation (2.62) is plotted in Figure 2.30 against the θ/τ ratio. This shows that, at smaller ratios, the process gain is slightly underestimated. At larger ratios, the approximation fails. Indeed, it becomes impossible to select a value

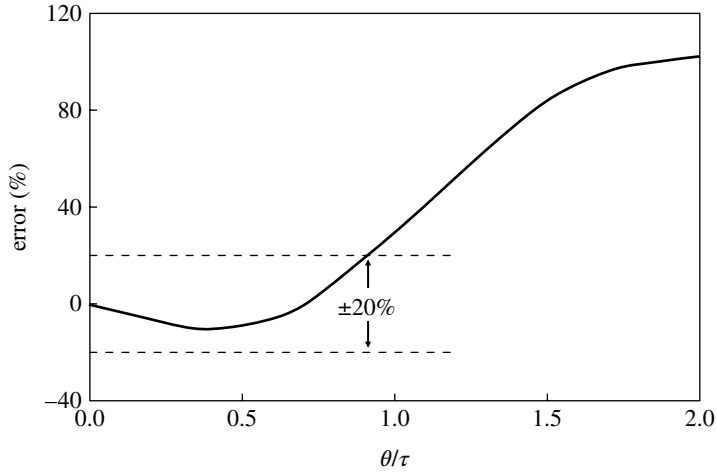

Figure 2.30 *Approximating self-regulating processes with integrating models*

for K'_p that gives acceptable control. Assuming we tolerate a 20% error in the estimate of K'_p, then we should not apply the approximation to processes where θ/τ is greater than 1. It is relatively rare for a process with a large lag also to have a large deadtime. The problem is that, while we can readily estimate θ by observing the beginning of the step test, we need to wait for steady state before we can estimate τ. So, unless we can be certain from our understanding of the process that θ is less than τ, we have to permit at least one step test to run to steady state.

Another approach to the problem is to simply conduct a series of step-tests, without waiting for the process to reach steady state between each step, and then applying the curve fitting technique described by Equation (2.16). The errors that such a method introduces reduce as the interval between steps is increased. Ideally we should ensure that at least one of the tests reaches steady state.

2.5 Other Types of Process

In addition to self-regulating and integrating processes, there are a range of others. There are processes which show a combination of these two types of behaviour. For example, steam header pressure generally shows integrating behaviour if boiler firing is changed. If there is a flow imbalance between steam production and steam demand, the header pressure will not reach a new steady state without intervention. However, as header pressure rises, more energy is required to generate a given mass of steam and the imbalance reduces. While the effect is not enough for the process to be self-regulating, the response will include some self-regulating behaviour.

Figure 2.31 shows another example. Instead of the temperature controller being mounted on a tray in the distillation column, it has been installed on the reboiler outlet. As the reboiler duty is increased, by increasing the flow of the heating fluid, the outlet temperature will increase. This will in turn cause the reboiler inlet temperature to increase – further increasing the outlet temperature which will then show integrating behaviour. However, the higher outlet temperature will result in increased vaporisation in the base of the column, removing some of the sensible heat as heat of vaporisation. Further, because of the reduction in temperature difference between the hot and cold side of the exchanger, the rate of heat transfer will decrease. This self-regulating effect will usually overcome the integrating behaviour and the process will reach a new steady state.

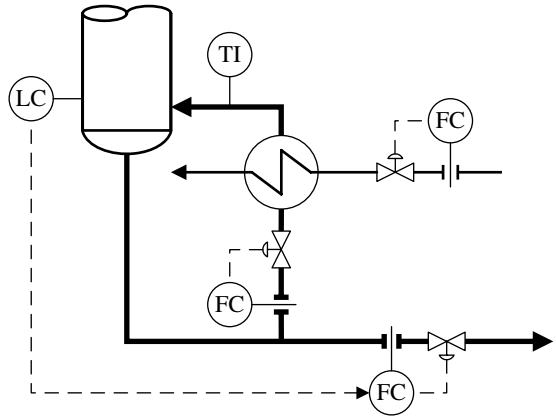

Figure 2.31 *Mixed integrating and self-regulating process*

The term *open-loop unstable* is also used to describe process behaviour. Some would apply it to any integrating process. But others would reserve it to describe inherently unstable processes such as exothermic reactors. Figure 2.32 shows the impact that increasing the reactor inlet temperature has on reactor outlet temperature. The additional conversion caused by the temperature increase generates additional heat which increases conversion further. It differs from other non-self-regulating processes in that the rate of change of PV increases over time. It is often described as a *runaway* response. Of course, the outlet temperature will eventually reach a new steady state when all the reactants are consumed; however, this may be well above the maximum permitted.

The term open-loop unstable can also be applied to controllers that have saturated. This means that the controller output has reached either its minimum or maximum output but not eliminated the deviation between PV and SP. The term can also be applied to a controller using a discontinuous on-stream analyser that fails. Such analysers continue to transmit the last measurement until a new one is obtained. If, as a result of analyser failure, no new measurement is transmitted, then the controller no longer has *feedback*.

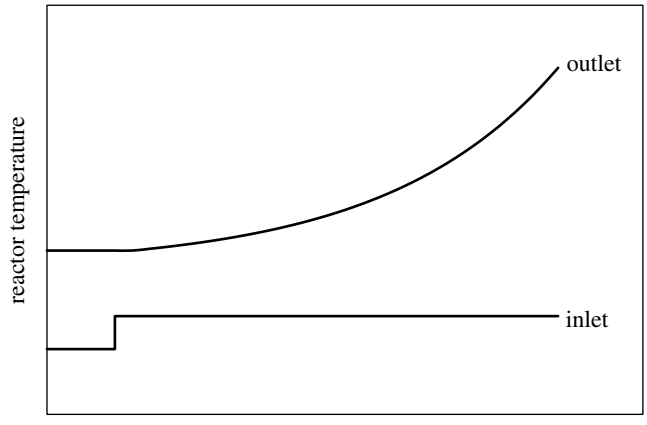

Figure 2.32 *Exothermic reactor*

2.6 Robustness

For a controller to be *robust* it must perform well over the normal variation of process dynamics. Dynamics are rarely constant and it is important to assess how much they might vary before finalising controller design.

Dynamics vary due to a number of reasons. The process may be inherently nonlinear so that, as process conditions vary, a controller tuned for one set of conditions may not work well under others. This is illustrated by Figure 2.33. A step test performed between points A and B would give a process gain of about 1.2, while one performed between points C and D would give a value of about 0.4. As a guideline, linear controllers are reasonably robust provided the process gain stays within ±20% of the value used to design the controller. In our example an average gain of 0.8 could be used but the variation would be ±50%. This would require a modified approach to controller design, such as the inclusion of some linearising function. It is important therefore that we conduct plant tests over the whole range of conditions under which the controller will be expected to operate.

A common oversight is not taking account of the fact that process dynamics vary with feed rate. Consider our example of a fired heater. If it is in a non-vaporising service, we can write the heat balance

$$F_{feed}.c_p\left(T - T_{inlet}\right) = F.NHV.\eta \tag{2.63}$$

On the feed side, F_{feed} is the flow rate to the heater, c_p is the specific heat, T is the outlet temperature and T_{inlet} is the inlet temperature. On the fuel side, F is the flow of fuel, NHV the net heating value (calorific value) and η the heater efficiency. Rearranging we get

$$T = \frac{F.NHV.\eta}{F_{feed}.c_p} + T_{inlet} \tag{2.64}$$

Differentiating:

$$K_p = \frac{dT}{dF} = \frac{NHV.\eta}{F_{feed}.c_p} \tag{2.65}$$

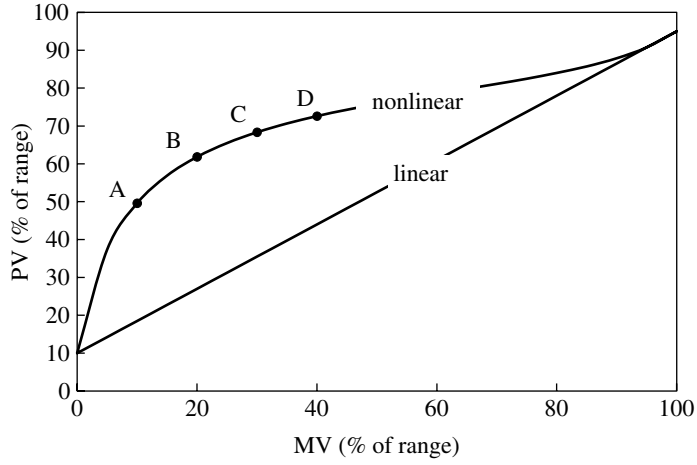

Figure 2.33 *Nonlinear process*

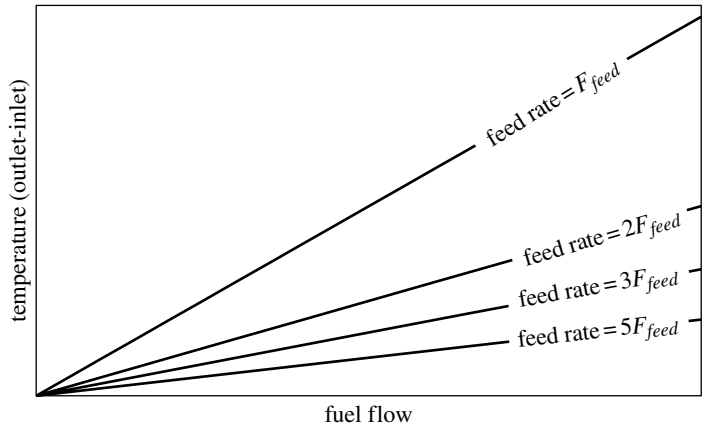

Figure 2.34 *Variation of process gain with feed rate*

While the process gain is sensitive to operating conditions, such as *NHV*, η and c_p, of most concern is its sensitivity to feed flow rate. In fact it is inversely proportional to feed rate. A little thought would have predicted this. Making the same increase in fuel at a higher feed rate would result in a smaller temperature increase because there is more feed to heat. Figure 2.34 shows how the relationship, between the rise in temperature across the heater and fuel flow, varies with feed rate. So, for example, doubling the feed rate halves the slope of the line. Some might describe the behaviour as nonlinear, using the term for any process in which the process gain is variable. Strictly, this is a linear process; changing feed rate clearly affects the process gain but behaviour remains linear at a given feed rate.

This effect is not unique to fired heaters; all process gains on a self-regulating process will vary with feed rate. Given that we tolerate ±20% variation in process gain, we can therefore tolerate ±20% variation in feed rate. Assuming a reference feed rate of 100, our controller will work reasonably well for feed rates between 80 and 120. The *turndown ratio* of a process is defined as the maximum feed rate divided by the minimum. We can see that, if this value exceeds 1.5 (120/80), the performance of almost all the controllers on the process will degrade noticeably as the minimum or maximum feed rate is approached. Fortunately, most processes have turndown ratios less than 1.5. Providing the controllers are tuned for the average feed rate, their performance should be acceptable. If this is not the case then techniques such as ratio feedforward, described in Chapter 6, can be applied. Of course, there are many other potential causes of variation in process gain. In general, if it is known to vary from a lower value of $(K_p)_1$ to a higher value of $(K_p)_2$, the average of these values can be used to design the controller – provided that the ratio of $(K_p)_2$ to $(K_p)_1$ is less than 1.5. If the variation is greater than this value, then adaptive tuning, as described in Chapter 3, should be applied.

Feed flow rate may also affect process deadtime. If the prime cause of deadtime is transport delay then an increase in flow will cause the residence time to reduce and so decrease deadtime. At worst, deadtime may be inversely proportional to feed rate. Figure 2.35 shows how control of a process degrades if controller tuning is not changed as dynamics change. We will show later that *integral over time of absolute error (ITAE)* is a measure of how poorly a process is controlled. Acceptable ±20% changes in process gain cause approximately a 20% increase in ITAE. However, for deadtime, controllers are far more sensitive to an increase than a decrease. Rather than design for the average, a value should be chosen so that variation

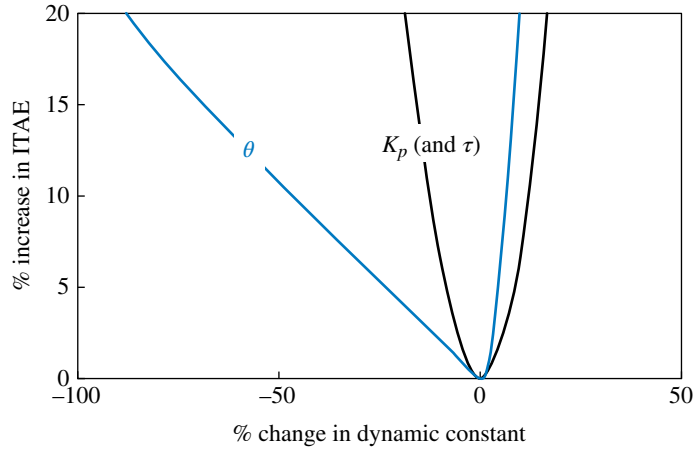

Figure 2.35 *Effect of process dynamics on controller performance*

does not exceed –90% to +10%. For example, if the process deadtime is known to vary between θ_1 and θ_2 (where $\theta_1 < \theta_2$), the deadtime used to design the controller would be given by

$$\theta = 0.1\theta_1 + 0.9\theta_2 \tag{2.66}$$

The variation in deadtime would then be acceptable provided θ_2/θ_1 is less than 11. Techniques for accommodating variation in excess of this are covered in Chapter 7.

Feed rate generally has little effect on process lag - although Equation (2.7) would appear to suggest otherwise. However, this only applies when there is complete mixing. In general, only in relatively small sections of most processes does this occur. But lag is often sensitive to the inventory of the process. For example, the lag caused by a vessel will change depending on the level of liquid in the vessel – as shown by Equation (2.7). Changes in vessel inlet temperature or composition will be more slowly detected at the vessel outlet if the level is high. Whether this is significant will depend on a number of factors. There are likely to be other sources of lag which, when added to that caused by the vessel, reduce the impact of inventory changes. Similarly, although the indicated level in the vessel may appear to change a great deal, it is unlikely that the level gauge operates over the full height of the vessel. A change in level from an indicated 10% to 90% would not mean that there is a nine-fold increase in liquid volume. However, a check should be made if averaging level control (as described in Chapter 4) is used – since this can permit large sustained changes in inventory. As Figure 2.35 shows, variations in process lag can be dealt with in the same way as process gain, i.e. by using the average of τ_1 and τ_2, provided that τ_2/τ_1 is less than 1.5.

If applying a curve fitting technique, as described earlier in this chapter, to identify process dynamics from a series of step tests then it may not be practical to determine how the dynamics vary between tests. To determine whether the process is sufficiently linear, it is more usual to check that the resulting model shows a good correlation between the PV and MV using, for example, the *Pearson R* correlation coefficient defined in Chapter 14. A value less than 0.95 might indicate nonlinearity. If the predicted and real process responses are compared, a nonlinear process would be apparent if increases are over-predicted and decreases under-predicted (or vice versa). Once the process deadtime and lag have been estimated, plotting the PV against the delayed and lagged MV will show the steady-state relationship. Using this to develop a linearisation function, applying it will enable the dynamics to be more accurately identified. However,

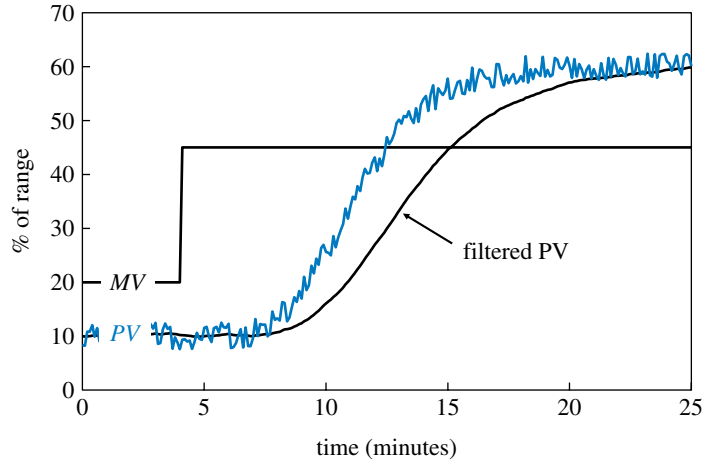

Figure 2.36 *Effect of filter on process dynamics*

the poor correlation might also be as a result of other process disturbances occurring during step-testing or of measurement noise. Repeating the test with larger steps would improve the correlation if this is the case.

The addition of filtering, to deal with measurement noise, can also affect the process dynamics. Figure 2.36 shows the same plant test but with noise added to the PV. This noise has then been removed by the addition of a filter (as described in Chapter 5). The filter adds lag and, because it increases the order of the system, also increases the apparent deadtime. Adding a filter after a controller has been tuned is therefore inadvisable. Either the plant test should be repeated to identify the new dynamics or, if the model identification package permits it, the original test data may be used with the filter simulated in the package. It is very common for filters to be implemented unnecessarily. They are often added visually to smooth the trended measurement. But the main concern should be the impact they have on the final control element, e.g. the control valve. This is a function not only of the amplitude of measurement noise but also the gains through which it passes. These may be less than one and so attenuate the noise. Not all filtering is implemented in the DCS. Most transmitters include filters. Provided the filter constant is not changed then model identification will include the effect of the transmitter filter in the overall dynamics. However, if the filter in the transmitter is changed by a well-intentioned instrument technician unaware of its implications, this can cause degradation in controller performance.

We will show later that controllers can be tuned to respond more quickly as K_p and θ/τ reduce. If dynamics can vary from those obtained by plant testing, it is better that the controller becomes more sluggish than more oscillatory. It is therefore safer to base controller tuning on higher values of K_p and θ, and on a lower value of τ.

3

PID Algorithm

The PID (proportional, integral, derivative) algorithm has been around since the 1930s. While many DCS vendors have attempted to introduce other more effective algorithms it remains the foundation of almost all basic control applications.

The basic form of the algorithm is generally well covered by academic institutions. Its introduction here follows a similar approach but extends it to draw attention to some of the more practical issues. Importantly it also addresses the many modifications on offer in most DCS, many of which are undervalued by industry unaware of their advantages. This chapter also covers controller tuning in detail. Several commonly known published methods are included but mainly to draw attention to their limitations. An alternative, well-proven, technique is offered for the engineer to use.

3.1 Definitions

Before proceeding, we must ensure that we define the key terminology. In Chapter 2 we defined *PV* (the process variable that we wish to control) and *MV* (the manipulated variable). The reader should note that some texts use this abbreviation to mean 'measured value', i.e. what we call *PV*. We will also use *M* to represent the controller output, which will normally be the same as *MV*. To these definitions we have also added *SP* (i.e. the target for *PV*).

The error (*E*) is defined as the deviation from *SP* but its definition varies between DCS. Our definition is

$$E = PV - SP \tag{3.1}$$

Most texts and some systems define error as $SP - PV$. Misinterpreting the definition will result in the controller taking corrective action in the direction opposite to that it should, worsening the error and driving the control valve fully closed or fully open.

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

3.2 Proportional Action

The principle behind proportional control is to keep the controller output (M) in proportion to the error (E).

$$M = K_c.E + C \tag{3.2}$$

K_c is the *controller gain* (occasionally called the *controller sensitivity*) and is a tuning constant set by the control engineer. The term C is necessary since it is unlikely to be the case that zero error coincides with zero controller output. In some control systems the value of C may be adjusted by the process operator, in which case it is known as *manual reset*. Its purpose will be explained later in this section.

We have seen that the process gain (K_p) may be positive or negative but the controller gain (K_c) is always entered into the control system as an absolute value. The control algorithm therefore includes an additional engineer-defined parameter known as *action*. If set to *direct* the controller output will increase as the PV increases; if set to *reverse*, output decreases as PV increases. If we consider our fired heater example, we would want the controller to reduce the fuel rate if the temperature increases and so we would need to set the action to reverse. In other words, if the process gain is positive, the controller should be reverse acting; if the process gain is negative, it should be direct acting. This definition is consistent with the concept of *negative feedback* where the product of all the gain terms in a control loop (known as the *loop gain*) must be negative. If feedback is positive the controller will quickly *saturate*, moving the MV to its minimum or maximum value, with probably a disastrous effect on the PV. The definition is also consistent with that adopted by the ISA [4] but is not used by all DCS vendors and is rare in textbooks. Some base the action on increasing E, rather than PV. If they also define error as $SP - PV$, then our heater temperature controller would need to be configured as direct acting.

Confusion can arise if the controller is manipulating a control valve. Valves are configured to either fail open or fail closed on loss of signal – depending on which is less hazardous. The signal actually sent to a 'fail open' valve therefore needs to be reverse acting. Some texts take this into account when specifying the action of the controller. However, most DCS differentiate between the output from the controller, which is displayed to the operator, and what is sent to the valve. To the operator and the controller all outputs represent the fraction (or percentage) that the valve is open. Any reversal required is performed after this. Under these circumstances, valve action need not be taken into account when specifying controller action.

The controller as specified in Equation (3.2) is known as the *full position* form in that it generates the actual controller output. A more useful form is the *incremental* or *velocity* form which generates the change in controller output (ΔM). We can convert the controller to this form by considering two consecutive scans. If E_n is the current error and E_{n-1} is the error at the previous scan then

$$M_n = K_c.E_n + C \tag{3.3}$$

$$M_{n-1} = K_c.E_{n-1} + C \tag{3.4}$$

Subtracting gives

$$\Delta M = K_c\left(E_n - E_{n-1}\right) = K_c.\Delta E \tag{3.5}$$

The advantage of this version is, firstly, that it eliminates C which is usually not a constant and would require adjustment as process conditions vary. Secondly, the controller will have *bumpless initialisation*. When any controller is switched from manual to automatic mode it should cause no disturbance to the process. With the

full position version it would be necessary to first calculate C to ensure that M is equal to the current value of the MV. Since the velocity form generates increments it will always start from the current MV and therefore requires no special logic. It also results in bumpless tuning, so that tuning constants may be changed with the controller in automatic mode. Some full position controllers require the controller to be first switched to manual and, after re-tuning, returned to automatic to force re-initialisation.

Some systems require the *proportional band (PB)* rather than gain. This is defined as the percentage change in error required to move the output 100%. Conversion between the two is straightforward.

$$PB = \frac{100}{K_c} \tag{3.6}$$

While it will respond to changes in PV, the main purpose of proportional action is to generate a *proportional kick* whenever the SP is changed. If we assume PV is constant then from Equation (3.1)

$$\Delta E = -\Delta SP \tag{3.7}$$

And, substituting into Equation (3.5)

$$\Delta M = -K_c.\Delta SP \tag{3.8}$$

Remembering that for our fired heater example the controller is reverse acting, the controller will thus make a step increase to fuel flow proportional to the increase in temperature SP. This is a one-off change because ΔSP will be zero for future scans until another change is made to SP. The response is shown in Figure 3.1. In this case K_c has been set to 2.

Of course, increasing the fuel will cause the temperature to rise and reduce the error – so the controller output will only remain at this new value until the process deadtime has expired. The full trend is shown in Figure 3.2. This demonstrates the main limitation of proportional control in that an *offset* (sometimes called *droop*) will always exist at steady state. The PV will never reach SP except at initial conditions. Figure 3.2

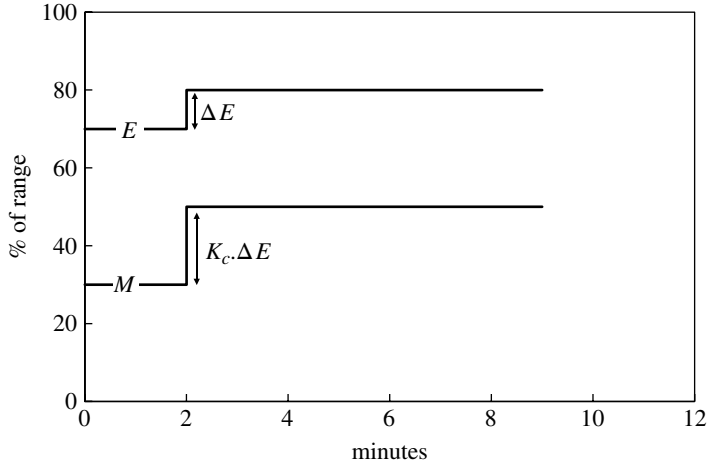

Figure 3.1 *Proportional kick*

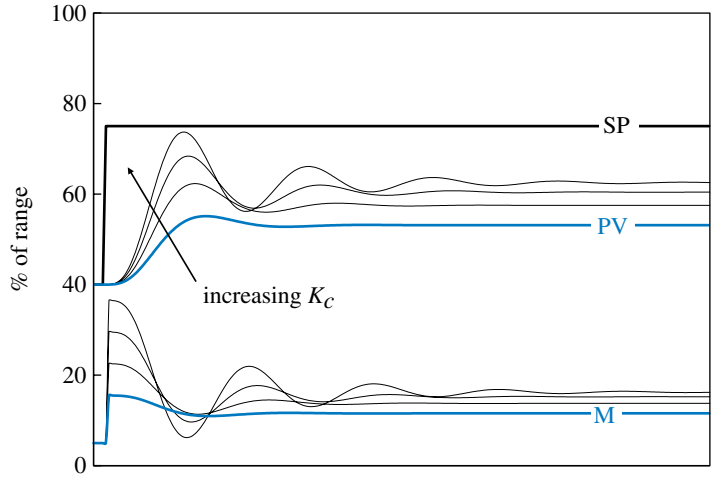

Figure 3.2 *Effect of changing K_c*

also shows that the offset can be reduced by increasing K_c but with increasing oscillatory behaviour. We will show that these oscillations, on any process, become unstable before offset can be reduced to zero.

Figure 3.3 represents a control system commonly found in the home. Found in lavatory cisterns and header tanks, it provides basic control of level. It operates by a float in effect measuring the deviation from the target level and a valve which is opened in proportion to the position of the float. It is a proportional-only controller. So why does it not exhibit offset? This is because it is not a continuous process. However, should it develop a continuous leak (flow = f), in order to maintain steady state the controller would have to maintain an inlet flow of f. The inlet flow can only be nonzero if the error is nonzero.

We can represent this mathematically. Before the leak develops the error is zero. When the process again reaches steady state the controller will have changed the inlet flow by f and the error will be E. By putting these values into Equation (3.5) we get

$$f = K_c.E \qquad \therefore E = \frac{f}{K_c} \tag{3.9}$$

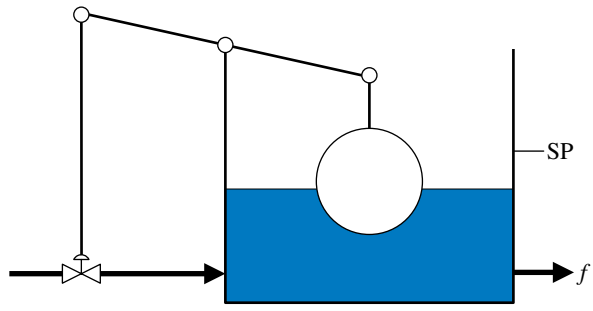

Figure 3.3 *Domestic level controller*

This confirms what we know, i.e. that offset can be reduced by increasing K_c but that it cannot be reduced to zero. In our cistern example this is equivalent to moving the pivot point closer to the float so that a smaller offset is required to achieve the same inlet flow. One might ask why not move the SP to account for the offset? One could, of course, and indeed we have the facility to do the equivalent by adjusting the manual reset term (C), as described in Equation (3.2). However, Equation (3.9) also shows us that the magnitude of the offset is proportional to the disturbance. Thus we would have to make this correction for virtually every disturbance and automation will have achieved little.

In general, offset is predictable provided the process gain is known. If we assume that the process gain is positive, we need a reverse acting controller. From Equation (3.5)

$$\Delta MV = -K_c.\Delta E = K_c\left(\Delta SP - \Delta PV\right) \tag{3.10}$$

By definition, for a self-regulating process

$$K_p = \frac{\Delta PV}{\Delta MV} \tag{3.11}$$

Combining these equations to eliminate ΔMV gives

$$\Delta PV = \frac{K_p.K_c}{1 + K_p.K_c}\Delta SP \tag{3.12}$$

The change in offset (ΔE) is therefore

$$\Delta E = \Delta PV - \Delta SP = \left(\frac{K_p.K_c}{1 + K_p.K_c} - 1\right)\Delta SP = \frac{-\Delta SP}{1 + K_p.K_c} \tag{3.13}$$

We showed previously that the concept of negative feedback means that $K_p.K_c$ is always negative. We have taken account of this by configuring Equation (3.10) as a reverse acting controller. We must therefore use the absolute value of $K_p.K_c$ in Equation (3.13).

Equation (3.13) forms the basis of some techniques which aim to identify process dynamics from closed loop testing. To re-tune an existing controller its integral and derivative actions are removed. The controller is switched to auto and its SP stepped. Since the error prior to the change is zero, the change in offset (ΔE) will also be the resulting offset (E). Remembering that E and ΔSP will have opposite signs, K_p can be obtained by rearranging Equation (3.13).

$$K_p = -\frac{\Delta SP + E}{K_c E} \tag{3.14}$$

We could, however, determine K_p as usual from observing the steady-state change (ΔPV) compared to the change made to the controller output (ΔMV). Indeed, combining Equations (3.14) and (3.10) gives Equation (3.11). We have simply used the controller to perform a step test. Of course, we also need estimates of θ and τ. These can be obtained by choosing K_c to give a slightly oscillatory response and measuring how this decays. However, because we covered several superior methods in Chapter 2, we will omit the detail of how this is done.

An offset can also arise from a *load* change. Load changes are process disturbances which cause the PV to move away from the SP. (In some industries, the term *regulator problem* is used to describe load changes, while *servo problem* is used to describe SP changes.) If the load is caused by a change of ΔDV in a disturbance variable then a similar analysis shows that offset is given by

$$\Delta PV = \frac{\left(K_p\right)_d \Delta DV}{1 + K_p K_c} \tag{3.15}$$

$(K_p)_d$ is the process gain between the PV and the DV.

Some texts state that an offset does not occur on an integrating process. We can demonstrate this is true for SP changes but it is not the case for load changes. For an integrating process

$$\frac{dPV}{dt} = K_p MV \tag{3.16}$$

But MV is set by the (proportional-only) controller

$$MV = K_c \left(PV - SP\right) \tag{3.17}$$

So

$$\frac{dPV}{dt} = K_p K_c \left(PV - SP\right) \tag{3.18}$$

At steady state

$$\frac{dPV}{dt} = 0 \tag{3.19}$$

So

$$PV = SP \tag{3.20}$$

For load changes Equation (3.16) becomes

$$\frac{dPV}{dt} = K_p MV + \left(K_p\right)_d DV \tag{3.21}$$

Applying the same approach we can show that the offset will not be zero.

$$PV - SP = \frac{-\left(K_p\right)_d DV}{K_p K_c} \tag{3.22}$$

Specifically for the level controller described above, if the DV is the outlet flow and the MV the inlet flow then, if the two flows have the same instrument range, $(K_p)_d$ will be numerically the same as K_p – although opposite in sign. Replacing DV with f, Equation (3.22) then becomes Equation (3.9).

One could argue that a proportional-only controller can be used for the secondary of a cascade because the primary will deal with the problem of offset by adjusting its SP. However, this undermines the main advantage of cascade control. We want the secondary to recover as quickly as possible from a process disturbance so that the disturbance to the primary is minimised.

It is not to say, however, that proportional-only control should never be used. There are situations where offset is acceptable (such as in some level controllers as described in Chapter 4). However, in most situations we need the PV always to reach the SP.

3.3 Integral Action

The main purpose of integral action is to eliminate offset. Sometimes called *reset action*, it continues to change the controller output for as long as an error exists. It does this by making the rate of change of output proportional to the error, i.e.

$$\frac{dM}{dt} = \frac{K_c}{T_i} E \tag{3.23}$$

T_i is known as the integral time and is the means by which the engineer can dictate how much integral action is taken. Equation (3.23) is already in the velocity form, integrating gives us the form that gives the action its name.

$$M = \frac{K_c}{T_i} \int E.dt \tag{3.24}$$

Converting Equation (3.24) to its discrete form (where *ts* is the controller scan interval) gives

$$M_n = \frac{K_c}{T_i} \sum_{j=0}^{n} E_j ts \tag{3.25}$$

While proportional action is based on the current error, we can see that integral action is based on (the sum of) past errors. Converting Equation (3.23) to the incremental form gives

$$\frac{\Delta M}{ts} = \frac{K_c}{T_i} E_n \tag{3.26}$$

In most systems the controller scan interval is expressed in seconds, while tuning constants such as T_i are often in minutes. Equation (3.26) and the others that follow therefore should then include a factor of 60 for conversion to consistent units. For simplicity we have omitted it.

Combining with Equation (3.5) gives proportional plus integral (PI) control

$$\Delta M = K_c \left[\left(E_n - E_{n-1} \right) + \frac{ts}{T_i} E_n \right] \tag{3.27}$$

The effect of the addition of integral action is shown in Figure 3.4. K_c has been reduced to a value of 1. The response shows that, for a constant error, the rate of change of output is constant. The change made by

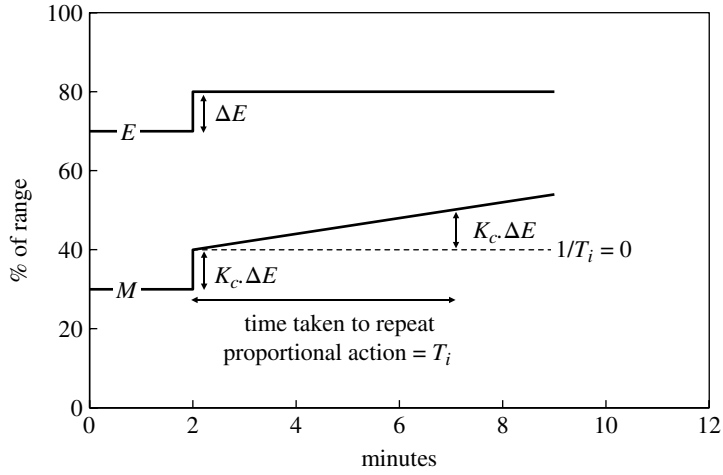

Figure 3.4 *Proportional plus integral action*

integral action will eventually match that of the initial proportional action. The time taken to 'repeat' the proportional action is T_i. In this example T_i is about 5 minutes. In many DCS T_i will have the units of minutes, but some systems use hours or seconds. Others define the tuning constant in *repeats per minute*, i.e. the reciprocal of T_i as we have defined it. The advantage of this is that, should more integral action be required, the engineer would increase the tuning constant. In the form of the algorithm we are using higher values of T_i give less integral action. We therefore have to be careful with the use of zero as a tuning constant. Fortunately most systems recognise this as a special case and disable integral action, rather than attempt to make an infinite change.

Again the trend in Figure 3.4 is only valid until the deadtime expires, after which the behaviour will be as shown in Figure 3.5. Even a very small amount of integral action will eliminate offset. Attempting to

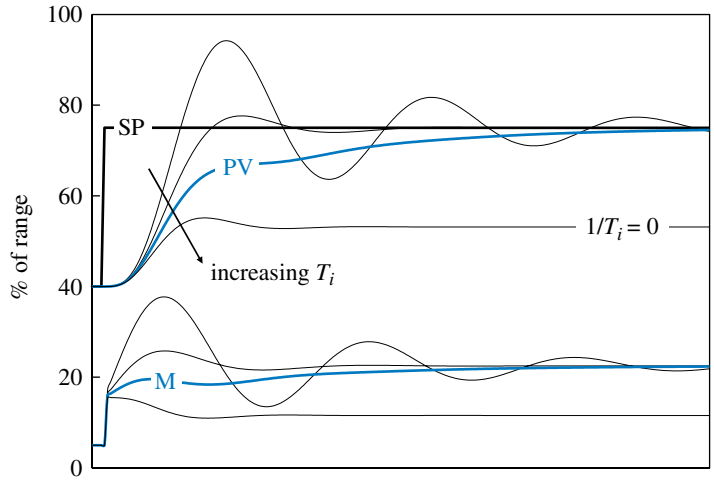

Figure 3.5 *Effect of changing T_i*

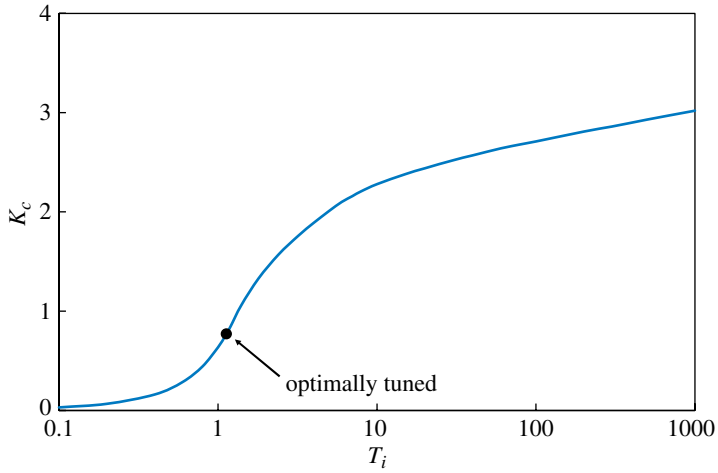

Figure 3.6 *Effect of integral action on optimum controller gain*

remove it too quickly will, as with any control action, cause oscillatory behaviour. However, this can be compensated for by reducing K_c. Optimum controller performance is a trade-off between proportional and integral action. For an arbitrarily chosen process ($K_p = 1$, $\theta = \tau = 1$ minute), Figure 3.6 illustrates how K_c should be adjusted as integral action is added to a P-only controller. The optimum K_c in a PI controller is substantially less than that in a P-only controller ($T_i \to \infty$).

3.4 Derivative Action

For most situations a PI controller is adequate. Indeed, many engineers will elect not to include derivative action to simplify tuning the controller by trial-and-error. A two-dimensional search for optimum parameters is considerably easier than a three-dimensional one. However, in most situations, the performance of even an optimally tuned PI controller can be substantially improved by the addition of derivative action.

Derivative action is intended to be anticipatory in nature; indeed, in older texts, it was called this. It anticipates by taking action if it detects a rapid change in error. The error may be very small (even zero) but, if changing quickly, will surely be large in the future. Derivative action attempts to prevent this by changing the output in proportion to the rate of change of error, i.e.

$$M = K_c.T_d \frac{dE}{dt} \tag{3.28}$$

T_d is known as the *derivative time* and is the means by which the engineer can dictate how much derivative action is taken. Converting Equation (3.28) to its discrete form, gives

$$M_n = K_c.T_d \frac{E_n - E_{n-1}}{ts} \tag{3.29}$$

Writing it for the previous scan interval

$$M_{n-1} = K_c.T_d \frac{E_{n-1} - E_{n-2}}{ts} \tag{3.30}$$

And subtracting

$$\Delta M = \frac{K_c.T_d}{ts}\left(E_n - 2E_{n-1} + E_{n-2}\right) \tag{3.31}$$

In order to demonstrate the effect of derivative action we will formulate a proportional plus derivative (PD) controller. This probably has no practical application but including integral action would make the trends very difficult to interpret. Combining Equations (3.5) and (3.31) gives

$$\Delta M = K_c\left[\left(E_n - E_{n-1}\right) + \frac{T_d}{ts}\left(E_n - 2E_{n-1} + E_{n-2}\right)\right] \tag{3.32}$$

Figure 3.7 shows the response of this controller, again up to the point where the deadtime expires. This time we have not made a step change to the SP, instead it has been ramped. The initial step change in the output is not then the result of proportional action but the derivative action responding to the change, from zero, in the rate of change of error. The subsequent ramping of the output is due to the proportional action responding to the ramping error. The proportional action will eventually change the output by the same amount as the initial derivative action. The time taken for this is T_d which, like T_i, can be expressed in units such as minutes or repeats per minute, depending on the DCS.

Also shown in Figure 3.7 is what the controller response would be without derivative action, i.e. proportional-only. It can be seen that derivative action takes action immediately and that the proportional action takes T_d minutes to do. In effect it has anticipated the need for corrective action, even though the error was zero at the time. We have seen that proportional action is based on current error and integral on past errors.

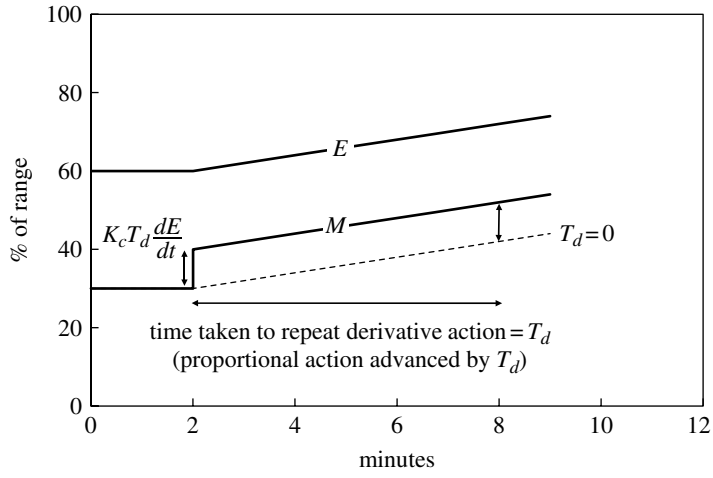

Figure 3.7 *Proportional plus derivative action*

Derivative is effectively based on future error. This anticipatory nature is beneficial if the process deadtime is large; it helps compensate for the delay between the change in PV and the cause of the disturbance.

Processes with a long deadtime, more specifically processes with a large θ/τ ratio, are relatively few. Thus most controllers, when responding to a change in SP, do not obviously benefit from the addition of derivative action. Indeed, if the θ/τ ratio is small, instability can be caused by relatively small amounts of derivative action. However, we will demonstrate later that derivative action permits K_c to be increased and so can be very useful in speeding the recovery from a load change – even if the θ/τ ratio is close to zero.

It is often said that derivative action should only be used in temperature controllers. It is true that temperatures, such as those on the outlet of a fired heater and on distillation column trays, will often exhibit significantly more deadtime than measurements such as flow, level and pressure. However this is not universally the case, as illustrated in Figure 3.8. Manipulating the bypass of the stream on which we wish to install a temperature controller, in this case around the tube side of the exchanger, will provide an almost immediate response. Indeed, if accurate control of temperature is a priority, this would be preferred to the alternative configuration of bypassing the shell side.

While there are temperatures with a very short deadtime there will be other measurements that, under certain circumstances, show a long deadtime. In Chapter 4 we include a level control configuration that is likely to benefit from derivative action. In Chapter 7 we describe a composition control strategy with a very large θ/τ ratio.

The full PID equation that we have developed so far is thus

$$\Delta M = K_c \left[\left(E_n - E_{n-1} \right) + \frac{ts}{T_i} E_n + \frac{T_d}{ts} \left(E_n - 2E_{n-1} + E_{n-2} \right) \right] \tag{3.33}$$

This form of the equation, however, exhibits a problem known as *derivative spike*. Consider how the derivative action responds to a change in SP. If, before the change, the process is at steady state and at SP then

$$E_{n-1} = E_{n-2} = 0 \tag{3.34}$$

The change will introduce an error (E) and so the change in output due to the derivative action will be

$$\Delta M = K_c \frac{T_d}{ts} E \tag{3.35}$$

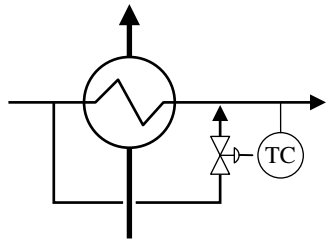

Figure 3.8 *Temperature control without deadtime*

Assuming the process deadtime is longer than the controller scan interval then, at the next scan, the PV will not yet have responded to this change and so both E_n and E_{n-1} will now have the value E. The derivative action will then be a change of the same magnitude but opposite in direction, i.e.

$$\Delta M = -K_c \frac{T_d}{ts} E \tag{3.36}$$

Until the process deadtime expires the values of E_n, E_{n-1} and E_{n-2} will all be E and so the derivative action will be zero. Bearing in mind that T_d will be of the order of minutes and ts in seconds the magnitude of ΔM is likely to be large, possibly even full scale, and is likely to cause a noticeable process upset. Derivative action is not intended to respond to SP changes. Remembering that we have defined error as $PV - SP$, then

$$E_n - 2E_{n-1} + E_{n-2} = PV_n - 2PV_{n-1} + PV_{n-2} - \left(SP_n - 2SP_{n-1} + SP_{n-2}\right) \tag{3.37}$$

If there are no SP changes, then

$$E_n - 2E_{n-1} + E_{n-2} = PV_n - 2PV_{n-1} + PV_{n-2} \tag{3.38}$$

And we can rewrite the PID controller as

$$\Delta M = K_c \left[\left(E_n - E_{n-1}\right) + \frac{ts}{T_i} E_n + \frac{T_d}{ts} \left(PV_n - 2PV_{n-1} + PV_{n-2}\right) \right] \tag{3.39}$$

This *derivative-on-PV* version (sometimes also described as PI-D) will no longer cause a derivative spike when there is a change in SP. But the response of derivative action to load changes is unaffected. In many DCS this modification is standard. Others offer both this and the *derivative-on-error* versions as options. As we will see later, it is common for this algorithm to include some form of filtering to reduce the impact of the spike but, even with this in place, there is no reason why the engineer should ever use the derivative-on-error version if the derivative-on-PV version is available.

While this modified algorithm deals with the problem of a spike resulting from a SP change, it will still produce spikes if there are steps in the PV. These can result if the measurement is discontinuous. The most common example is some types of on-stream analysers, such as chromatographs. The *sample-and-hold* technique these employ will exhibit a staircase trend as the PV changes. Each step in the staircase will generate a spike. This is a particular issue because analysers tend to be a significant contributor to deadtime and thus the composition controller would benefit from the use of derivative action. This problem is addressed in Chapter 7.

However, the problem can also arise from the use of digital field transmitters, even if the analog-to-digital conversion is done to a high resolution, say, to 0.1% of range. The resulting 0.1% steps as the PV changes can be amplified by one or two orders of magnitude by the derivative action. Care should therefore be taken in the selection of such transmitters if they are to be installed in situations where derivative action would be beneficial.

Figure 3.9 shows the performance of the full PID controller as described in Equation (3.39). As might be expected, the response becomes more oscillatory as T_d is increased. Perhaps more surprising is that reducing T_d also causes an oscillatory response. This is because the addition of derivative action permits an increase in controller gain, so the oscillation observed by removing the derivative action is caused by excessive proportional action.

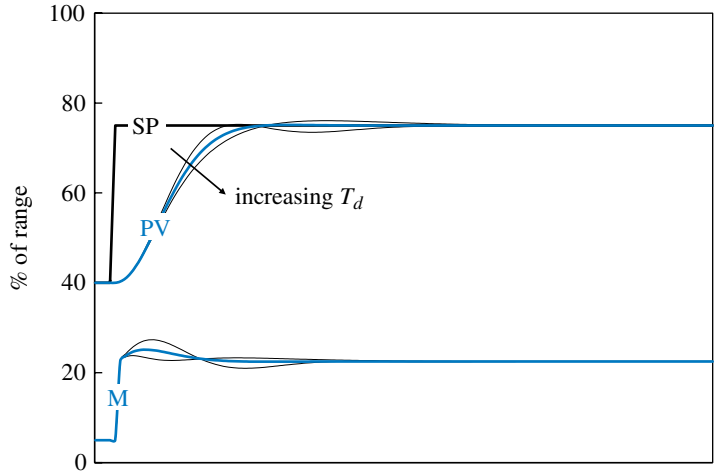

Figure 3.9 *Effect of changing T_d*

This interdependence means that we cannot simply add derivative action to a well-tuned PI controller. It will only be of benefit if all three tuning constants are optimized. Similarly, if we wished to remove derivative action from a controller, we should re-optimize the proportional and integral tuning. For the same arbitrarily chosen process ($K_p = 1, \theta = \tau = 1$ minute), Figure 3.10 illustrates how K_c should be adjusted as derivative action is added to a PI controller. Not shown, because the impact is relatively small, is the variation in T_i. K_c for an optimally tuned PID controller is, in this case, about 40% higher than that for an optimally tuned PI controller. Further increase in derivative action is possible but adds instability to the controller, ultimately requiring K_c to be reduced.

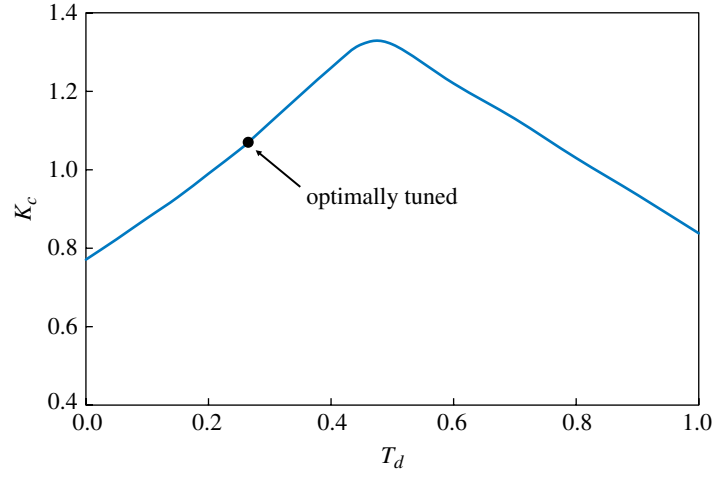

Figure 3.10 *Effect of derivative action on optimum controller gain*

If measurement noise is present then we need to be cautious with the application of derivative action. While the amplitude of the noise may be very small, it will cause a high rate of change of the PV. Derivative action will therefore amplify this. We can illustrate this by assuming the noise is a sinusoidal signal with amplitude A and frequency f. If the process is steady and at SP then the error (E) will be given by

$$E = A\sin\left(2\pi f.t\right) \tag{3.40}$$

Differentiating with respect to time (t) shows that the change in output caused by the derivative action will also be sinusoidal (shifted in phase), with the same frequency, but amplified by $2\pi f$. The higher the frequency, the greater will be the amplification.

$$\frac{dE}{dt} = A.2\pi f\,\cos\left(2\pi f.t\right) = A.2\pi f\,\sin\left(2\pi f.t + \frac{\pi}{2}\right) \tag{3.41}$$

The possible presence of noise is perhaps another reason why there may be a reluctance to use derivative action. However modern DCS provide a range of filtering techniques which can permit it to be applied effectively. These are covered in Chapter 5.

Figure 3.11 shows the benefit of including derivative action. The open loop response was produced by making a step change to the MV of the same magnitude as that ultimately made by the controllers. By applying the methods detailed earlier in this chapter, the reader can confirm that this is a long deadtime process with a θ/τ ratio of about 2.7. The closed loop responses were placed by overlaying their trends so that the change in SP is at the same point in time as the start of the open loop test. With such a process an optimally tuned PID controller will outperform an optimally tuned PI controller by reaching SP in about 30% less time. Given that it is impossible for any controller to reach SP before the deadtime has elapsed, this is a substantial improvement. As a general rule, derivative action will make a noticeable improvement to the response to SP changes if the θ/τ ratio is greater than 0.5. We will show later that it benefits the response to load changes at much lower values of θ/τ.

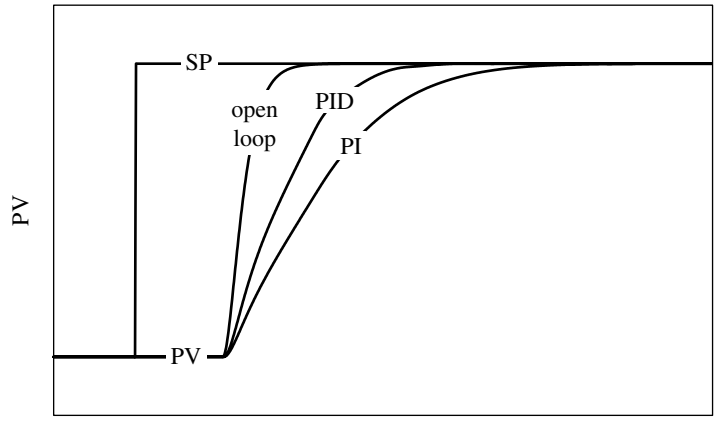

Figure 3.11 *Benefit of derivative action*

3.5 Versions of Control Algorithm

Before embarking on tuning the controller it is important that we understand the exact form of the algorithm. Different versions are in common use for two reasons. Firstly, there is a variety of approaches taken by different DCS vendors in converting the equations written in analog form into their discrete version. Secondly, vendors have added a range of enhancements to the 'standard' controller.

Addressing the first of these issues we can write, by combining Equations (3.2), (3.24) and (3.28), the conventional analog *time domain* version of the algorithm.

$$M = K_c\left[E + \frac{1}{T_i}\int E.dt + T_d\frac{dE}{dt}\right] + C \tag{3.42}$$

To eliminate C we differentiate the equation

$$\frac{dM}{dt} = K_c\left[\frac{dE}{dt} + \frac{E}{T_i} + T_d\frac{d}{dt}\left(\frac{dE}{dt}\right)\right] \tag{3.43}$$

Approximating with finite differences we then get

$$\frac{\Delta M}{\Delta t} = K_c\left[\frac{\Delta E_n}{\Delta t} + \frac{E_n}{T_i} + \frac{T_d}{\Delta t}\left(\frac{\Delta E_n}{\Delta t} - \frac{\Delta E_{n-1}}{\Delta t}\right)\right] \tag{3.44}$$

Replacing Δt with the controller scan interval (ts) gives the version of the PID algorithm described by Equation (3.33) but another approach is to rewrite Equation (3.42) directly in its discrete form.

$$M_n = K_c\left[E_n + \frac{ts}{T_i}\sum_{j=0}^{n}E_j + \frac{T_d}{ts}\left(E_n - E_{n-1}\right)\right] + C \tag{3.45}$$

In doing so we have applied the *rectangular rule*, i.e. the integral is treated as a series of rectangles of width ts and height E_j. Since, unlike the incremental form, this algorithm does not inherently initialise bumplessly it must be designed to do so. This is achieved by determining what value should initially be used for the accumulated error in order for C to be zero. When not in automatic mode, the calculation performed by the controller becomes

$$\sum_{j=0}^{n}E_j = \frac{T_i}{ts}\left[\frac{M_n}{K_c} - E_n - \frac{T_d}{ts}\left(E_n - E_{n-1}\right)\right] \tag{3.46}$$

The output (M_n) is now the desired output. If the controller is directly manipulating a valve, M_n is the current valve position; if the primary of a cascade, M_n is the SP of the secondary. On switching to automatic, M will therefore not change. To achieve bumpless tuning a similar initialisation calculation must be performed at the next scan following the tuning calculation. By inserting the old and new tuning constants into Equation (3.46), the difference between the two results is the amount by which the accumulated error must be changed. Alternatively, by subtracting from Equation (3.45) the same equation written for the previous scan ($n-1$), we obtain the incremental algorithm described in Equation (3.33).

An alternative method is to apply the *trapezium rule*, where the integral is treated as a series of trapeziums.

$$M_n = K_c' \left[E_n + \frac{ts}{T_i'} \sum_{j=0}^{n} \frac{E_j + E_{j-1}}{2} + \frac{T_d'}{ts} \left(E_n - E_{n-1} \right) \right] + C \tag{3.47}$$

We can again design this controller to initialise bumplessly but, by subtracting the equation for the previous scan, we obtain (a slightly different version of) the velocity form of the controller.

$$\Delta M = K_c' \left[\left(E_n - E_{n-1} \right) + \frac{ts}{T_i'} \left(\frac{E_n + E_{n-1}}{2} \right) + \frac{T_d'}{ts} \left(E_n - 2E_{n-1} + E_{n-2} \right) \right] \tag{3.48}$$

This uses the average of the last two errors in the integral action rather than the latest value of the error. The algorithm will perform in exactly the same way, provided that the tuning is adjusted to take account of the change. By equating coefficients for E_n, E_{n-1} and E_{n-2}we can derive tuning constants for Equation (3.48) from those used in Equation (3.33).

$$K_c' = K_c \left[1 + \frac{ts}{2T_i} \right] \tag{3.49}$$

$$T_i' = T_i + \frac{ts}{2} \tag{3.50}$$

$$T_d' = T_d \left[\frac{2T_i}{2T_i + ts} \right] \tag{3.51}$$

Details of the Laplace form of the control equations are presented in Chapter 15. Without going into detail, for the 'standard' controller, it would be

$$M = K_c \left[1 + \frac{1}{T_i s} + T_d s \right] E \tag{3.52}$$

Again, without going into details, a more rigorous conversion using the z-transform gives the finite difference form

$$\Delta M = K_c' \left[\left(E_n - E_{n-1} \right) + \frac{ts}{T_i'} E_{n-1} + \frac{T_d'}{ts} \left(E_n - 2E_{n-1} + E_{n-2} \right) \right] \tag{3.53}$$

Again, only the integral term is affected – using E_{n-1} rather than E_n. Again, by equating coefficients we can show that

$$K_c' = K_c \left[1 + \frac{ts}{T_i} \right] \tag{3.54}$$

$$T_i' = T_i + ts \tag{3.55}$$

$$T_d' = T_d \left[\frac{T_i}{T_i + ts} \right] \tag{3.56}$$

Equations (3.49) to (3.51) and Equations (3.54) to (3.56) all show that, as *ts* tends towards zero, the tuning constants are the same for all three versions of the PID controller. This confirms that all three are good approximations to analog control. It also shows that, provided the tuning constants are large compared to the scan interval, the values required vary little between the algorithms. Tuning constants are generally of the order of minutes, while the scan interval is of the order of seconds, and so it will generally be the case that we do not need to know the precise form of the algorithm. This is somewhat fortunate with many DCS; the vendor will describe the algorithm in its analog form but not always divulge how it has been converted to its discrete form. However, if the process dynamics are very fast, resulting in tuning constants measured in a few seconds, then knowing the precise form of the algorithm becomes important.

3.6 Interactive PID Controller

There is a quite different version of the PID algorithm. Equation (3.42) can be represented diagrammatically as in Figure 3.12. To convert the box diagram to the equation, functions in parallel are additive, those in series are multiplicative.

In line with many control system vendors, we will describe this algorithm as *ideal*, using the term *parallel* to describe the algorithm illustrated in Figure 3.13. This version is truly *non-interactive* since each of the tuning constants can be adjusted with no effect on the other actions. In the case of the ideal algorithm, adjusting K_c will affect all three actions. The difference between the two is trivial since it is possible to easily set the tuning of the ideal version so that it performs identically. K_c is the same for both algorithms; only T_i and T_d require modification.

$$T_i = K_c T_i' \qquad \text{and} \qquad T_d = \frac{T_d'}{K_c} \tag{3.57}$$

However, the parallel algorithm is rarely found in DCS. The term ‘non-interactive’ is therefore often applied to the more common ‘ideal’ version.

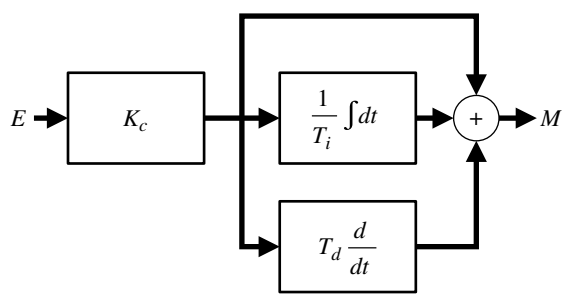

Figure 3.12 *Ideal PID*

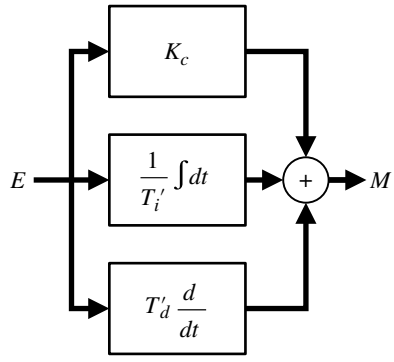

Figure 3.13 Parallel PID

The *series* version is more representative of the algorithm used by early pneumatic instruments and is retained, usually as an option, in some DCS. It arises from the way derivative action is added to the PI algorithm. Equation (3.58) defines the *projected error* that will exist at time T_d' if the current rate of change of error is maintained.

$$\hat{E} = E + T_d' \frac{dE}{dt} \tag{3.58}$$

We then use this value, instead of the current error, replacing E in the PI algorithm

$$M = K_c' \left[\hat{E} + \frac{1}{T_i'} \int \hat{E}.dt \right] = K_c' \left[E + T_d' \frac{dE}{dT} + \frac{1}{T_i'} \left(\int E.dt + T_d' E \right) \right] \tag{3.59}$$

Rearranging gives

$$M = K_c' \left[\left(1 + \frac{T_d'}{T_i'} \right) E + \frac{1}{T_i'} \int E.dt + T_d' \frac{dE}{dT} \right] \tag{3.60}$$

Figure 3.14 shows the block structure of this algorithm. Comparison with our ideal controller in Equation (3.42) shows that the proportional action depends not only on K_c' but also on T_i' and T_d'. Changing either

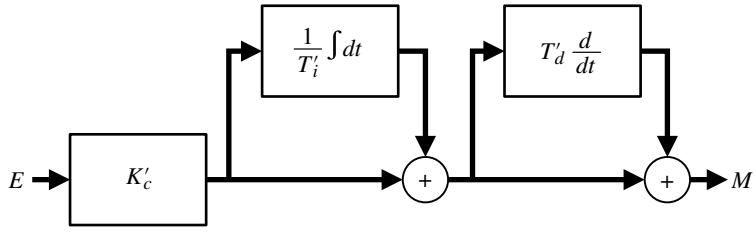

Figure 3.14 Series PID

the integral action or the derivative action will affect the proportional action – hence its 'interactive' title. Equation (3.60) may be rewritten.

$$M = K'_c \left(\frac{T'_i + T'_d}{T'_i} \right) \left[E + \frac{1}{T'_i + T'_d} \int E.dt + \frac{T'_d T'_i}{T'_i + T'_d} \frac{dE}{dT} \right] \tag{3.61}$$

And hence, to switch from the interactive to the ideal version therefore requires the tuning constants to be changed as follows.

$$K_c = K'_c \frac{T'_i + T'_d}{T'_i} \tag{3.62}$$

$$T_i = T'_i + T'_d \tag{3.63}$$

$$T_d = \frac{T'_d T'_i}{T'_i + T'_d} \tag{3.64}$$

By rearranging Equations (3.62) to (3.64) we can obtain the equations for converting ideal tuning to that for the interactive algorithm.

$$K'_c = \frac{K_c}{2} \left[1 + \sqrt{1 - 4 \frac{T_d}{T_i}} \right] \tag{3.65}$$

$$T'_i = \frac{T_i}{2} \left[1 + \sqrt{1 - 4 \frac{T_d}{T_i}} \right] \tag{3.66}$$

$$T'_d = \frac{2T_d}{\left[1 + \sqrt{1 - 4 \frac{T_d}{T_i}} \right]} \tag{3.67}$$

While the interactive and ideal algorithms can be tuned to give identical performance, this is subject to the restriction that $T_d < 0.25T_i$. Indeed, if no derivative action is required, the algorithms will perform identically without the need to change tuning. Combining Equations (3.66) and (3.67) we obtain

$$\frac{T'_d}{T'_i} = \frac{4T_d}{T_i \left(1 + \sqrt{1 - 4 \frac{T_d}{T_i}} \right)^2} \tag{3.68}$$

We will show later that T_d/T_i should be increased as the θ/τ ratio increases. Figure 3.15 shows a typical relationship for the ideal control algorithm. Applying Equation (3.68) allows us also to show how T'_d/T'_i

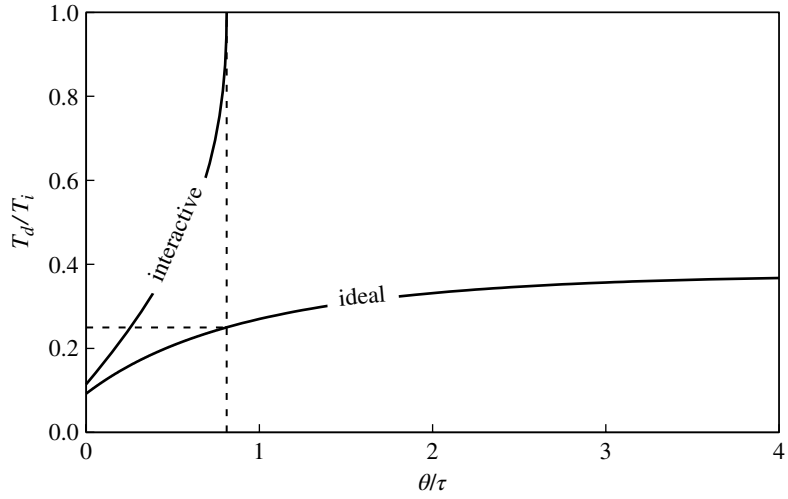

Figure 3.15 *Effect of algorithm choice on T_d/T_i*

varies for the interactive version. This demonstrates that the ratio is always higher than that required by the ideal version – indeed, reaching a value of 1 as T_d/T_i reaches 0.25. So, provided the θ/τ ratio is less than about 0.8, both algorithms can be tuned to give identical performance.

So that one algorithm can be adopted as the standard approach for all situations, there are several arguments for choosing the ideal algorithm. Firstly, it can be tuned to give better performance for processes with a large θ/τ ratio; these merit much more derivative action with T_d often approaching $0.4\,T_i$. Secondly, most DCS use the ideal algorithm; others give the option to use either. Finally, most published tuning methods are based on the ideal version.

In its digital form the interactive algorithm is

$$\Delta M = K_c\left[\left(1+\frac{T_d}{T_i}\right)\left(E_n - E_{n-1}\right)+\frac{ts}{T_i}E_n+\frac{T_d}{ts}\left(E_n - 2E_{n-1}+E_{n-2}\right)\right] \tag{3.69}$$

In Laplace form, it is written as

$$M = K_c\left[1+\frac{1}{T_i s}\right]\left[1+T_d s\right]E \tag{3.70}$$

But it is usually modified to

$$M = K_c\left[1+\frac{1}{T_i s}\right]\left[\frac{1+T_d s}{1+aT_d s}\right]E \tag{3.71}$$

Full details of Laplace transforms can be found in Chapter 15 but comparison with Equation (3.70) shows that an additional term has been introduced, i.e. $(1 + aT_d s)$. This introduces a lag into the controller (of time constant aT_d) that is intended to reduce the amplification of measurement noise by the derivative action. The algorithm can be written without using Laplace but the resulting complexity does not make

any more apparent the purpose of the modification. Although almost certainly not exactly how any vendor has implemented it, the closest discrete approximation is

$$\Delta M = e^{-ts/aT_d}\Delta M_{n-1} + K_c\left(P + I + D\right) \tag{3.72}$$

where

$$P = \left[1 + \left(1-a\right)\frac{T_d}{T_i}\right]\left[E_n - E_{n-1} - e^{-ts/aT_d}\left(E_{n-1} - E_{n-2}\right)\right] \tag{3.73}$$

$$I = \frac{ts}{T_i}\left[E_n - e^{-ts/aT_d}E_{n-1}\right] \tag{3.74}$$

$$D = \left(1-a\right)\frac{T_d}{ts}\left[1 - a\frac{T_d}{T_i}\right]\left[E_n - 2E_{n-1} + E_{n-2}\right] \tag{3.75}$$

Setting a to zero removes the filter, setting it to 1 will completely disable the derivative action. In some systems the value of a is configurable by the engineer. In most it is fixed, often at a value of 0.1. The reciprocal of a is known as the *derivative gain limit*. Its inclusion is of debatable value. As we shall show later, derivative action scanning at too high a frequency will greatly amplify the noise transmitted to the actuator. Under these circumstances the derivative filter will appear beneficial when perhaps a better approach would be to increase the scan interval – especially as this will also help unload the control system. If no noise is present and derivative action is required then we have to modify the controller tuning to take account of the inclusion of a. Few published tuning methods do this. More sophisticated tuning methods permit the value of a to be optimised along with K_c, T_i and T_d. Almost invariably, these methods show that the optimum value for a is zero. Further, the filtering is identical to that provided by the standard DCS filter (see Chapter 5). The DCS filter is generally adjustable by the engineer, whereas that in the control algorithm often is not. Indeed, values of a of 0.25 or less can have little impact on the transmission of noise through to the MV. Even if a is adjustable, its upper limit means that the filter time constant cannot be increased beyond T_d. If noise is an issue then an engineer-configurable filter is preferred. This strengthens the argument not to use the interactive version of the controller.

The inclusion of a also reduces the size of the derivative spike, as quantified by Equation (3.35). If a is 0.1, even if T_d/T_i is at its maximum value of 1, Equation (3.75) shows that the spike is reduced by only about 19%. The problem is far better resolved by the use of the derivative-on-PV algorithm.

Finally, the inclusion of the term a makes impossible the development of a simple formula that converts tuning for the ideal algorithm to that required for this version of the interactive algorithm – and vice versa. Instead we have to revert to obtaining the process dynamics and derive the tuning as we would for a new controller.

Some vendors have also introduced filtering into the ideal version as

$$M = K_c\left[1 + \frac{1}{T_i s} + \frac{T_d s}{1 + aT_d s}\right]E \tag{3.76}$$

Comparison with Equation (3.52) shows that the lag is only applied to the derivative action and thus does not force unnecessary filtering of the proportional and integral actions. We will describe in Chapter 5 how such a filter is formulated but the derivative action is based on the error modified as follows:

$$\hat{E}_n = e^{-ts/aT_d}\hat{E}_{n-1} + \left(1 - e^{-ts/aT_d}\right)E_n \tag{3.77}$$

The control algorithm, for example, that defined by Equation (3.33), then becomes

$$\Delta M = K_c\left[\left(E_n - E_{n-1}\right) + \frac{ts}{T_i}E_n + \frac{T_d}{ts}\left(\hat{E}_n - 2\hat{E}_{n-1} + \hat{E}_{n-2}\right)\right] \tag{3.78}$$

As we will see in Chapter 5, there are approximate alternatives to Equation (3.77) that avoid using the exponential function. Often adopted by the DCS vendor they will affect the behaviour of the resulting PID algorithm – so theoretically requiring different tuning. However, provided the process dynamics are much larger than *ts*, the impact is unlikely to be noticeable.

A more rigorous formulation of the algorithm can be derived along the lines of Equation (3.72), where

$$P = E_n - E_{n-1} - e^{-ts/aT_d}\left(E_{n-1} - E_{n-2}\right) \tag{3.79}$$

$$I = \frac{ts}{T_i}\left[E_n - e^{-ts/aT_d}E_{n-1}\right] \tag{3.80}$$

$$D = \frac{T_d}{ts}\left(1 - e^{-ts/aT_d}\right)\left[E_n - 2E_{n-1} + E_{n-2}\right] \tag{3.81}$$

This shows that setting *a* to zero gives the conventional PID algorithm as described by Equation (3.33). As *a* tends to infinity, the derivative action becomes negligible and the algorithm becomes the conventional PI controller, as described by Equation (3.27). However, as in the interactive version, *a* may still not be adjustable by the engineer and is not usually taken account of in published tuning methods. If not adjustable it is set at a value, typically too small to have a noticeable effect on the level of noise in the controller output. Its inclusion presents the same issues as those described for the interactive version.

3.7 Proportional-on-PV Controller

The most misunderstood and most underutilised version of the PID algorithm is the *proportional-on-PV* (or I-PD) type. Taking the algorithm developed as Equation (3.39) we modify the proportional action so that it is based on PV rather than error.

$$\Delta M = K_c\left[\left(PV_n - PV_{n-1}\right) + \frac{ts}{T_i}E_n + \frac{T_d}{ts}\left(PV_n - 2PV_{n-1} + PV_{n-2}\right)\right] \tag{3.82}$$

In the same way that changing the derivative action from using PV instead of error, this change will stop the proportional action responding to changes in SP. This would appear to undermine the main purpose of

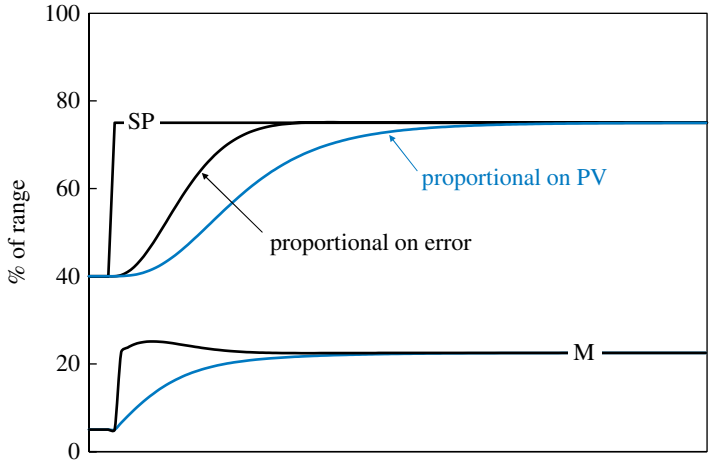

Figure 3.16 *P-on-PV algorithm*

proportional action by eliminating the proportional kick it would otherwise produce whenever the SP is changed. Indeed, only the integral action will now respond to the SP change, producing a much gentler ramping function. This can be seen in Figure 3.16 where a well-tuned proportional-on-error algorithm has had this modification made. The absence of the initial proportional kick can be seen on the trend of the MV and results in the PV taking much longer to reach its new SP.

Many believe therefore that this algorithm should be applied on processes where the MV should be adjusted slowly. However, if this performance were required, it could be achieved by tuning the more conventional proportional-on-error algorithm. Conversely it is important to recognise that the proportional-on-PV algorithm can be re-tuned to compensate for the lack of the proportional kick and so respond well to SP changes. This is illustrated in Figure 3.17.

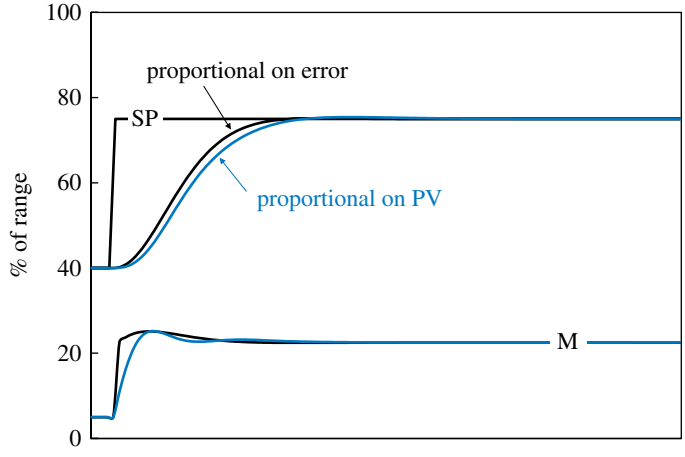

Figure 3.17 *P-on-PV algorithm retuned for SP change*

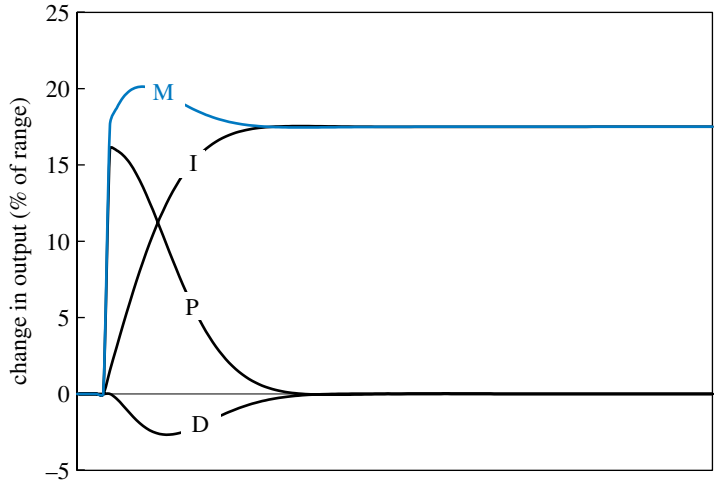

Figure 3.18 *Response to SP change (P-on-E algorithm)*

Figure 3.18 shows the behaviour of each part of the proportional-on-error control algorithm in response to the SP change above. The proportional kick is clear with the proportional part of the controller returning to zero as the error returns to zero. The derivative action is the greatest as the PV peaks, and so permits more proportional and integral action to be used. It too returns to zero as the rate of change of PV returns to zero.

Figure 3.19 shows the same disturbance but with the proportional-on-PV algorithm. Note that the vertical scale is much larger than that in Figure 3.18. As expected, there is no proportional kick and, since the action is now based on PV, the proportional part does not return to zero. On increasing the SP, the error will decrease but the PV increases. On our heater we require a reverse acting controller – explaining why the proportional part now reduces as the SP is increased. The integral action compensates for this so that there is a net increase in controller output. The derivative action behaves in almost the same way as in the proportional-on-error case, but the correction is larger because of the higher controller gain.

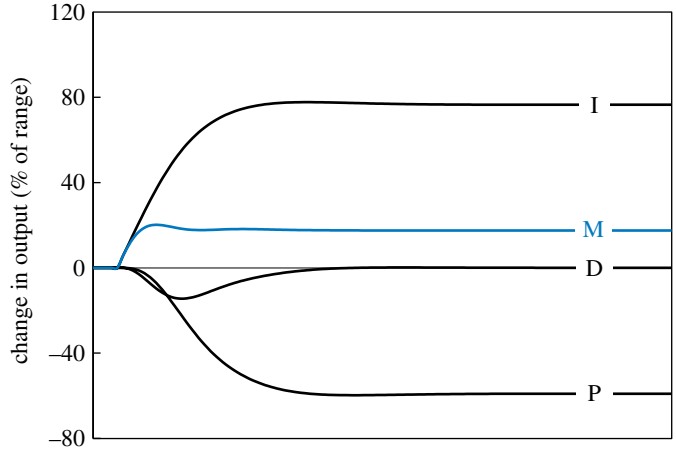

Figure 3.19 *Response to SP change (P-on-PV algorithm)*

Figure 3.17 shows that the performance of a retuned proportional-on-PV controller would be, on a real process, indistinguishable from the original proportional-on-error controller. Compensation for the loss of the proportional kick has been achieved mainly by substantially increasing the controller gain. This causes the integral action to ramp the MV much faster. But achieving similar performance begs the questions as to why the proportional-on-PV algorithm is included in most DCS and when it should be used.

Rather than consider SP changes, we should give more attention to load changes. While most tuning methods are designed to deliver *SP tracking*, we also need *disturbance rejection*. On our fired heater, for example, changes in feed flow rate or heater inlet temperature would cause such deviation from the SP. Most controllers experience many more load changes than SP changes. A heater outlet temperature controller may operate for days or weeks with no change to its SP. But it is likely to experience many process disturbances in the meantime.

Load changes impact the error differently to SP changes since their effect must pass through the process and is subject to the process lag. Rather than in the case of a SP change, when the error changes as a step, it will accumulate more gradually. Figure 3.20 shows the performance of the two controllers, both tuned for SP changes, subjected to a load change. The change could, for example, be an increase in feed flow rate. The open loop trend shows what would happen with no temperature control in place.

Switching the algorithm between proportional-on-error and proportional-on-PV has no effect on the way it responds to load changes. The difference we see is due to the difference in tuning. The more tightly tuned algorithm deviates from SP by less than half and the duration of the upset is also halved. This opportunity for substantial improvement is often overlooked by control engineers. Preoccupied with tuning controllers for SP changes, they rarely appreciate how much faster the controller can be made to react to process disturbances.

Figure 3.21 shows the breakdown of the control action for the load tuning case above. Since the SP is constant, the response would the same for both the proportional-on-error and proportional-on-PV algorithms.

It is important to recognise that the process cannot benefit from the tuning change if applied to the proportional-on-error algorithm. The effect of doing so is shown in Figure 3.22. Even if SP changes are rare, when they are made, the controller will now react far too quickly. In our example the MV has overshot its steady-state change by over 200%. In our fired heater example, this would likely cause a serious upset to fuel combustion.

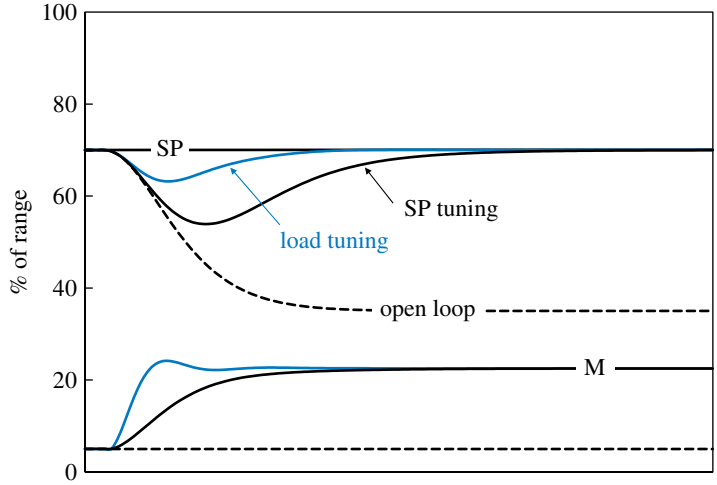

Figure 3.20 *Response to a load change*

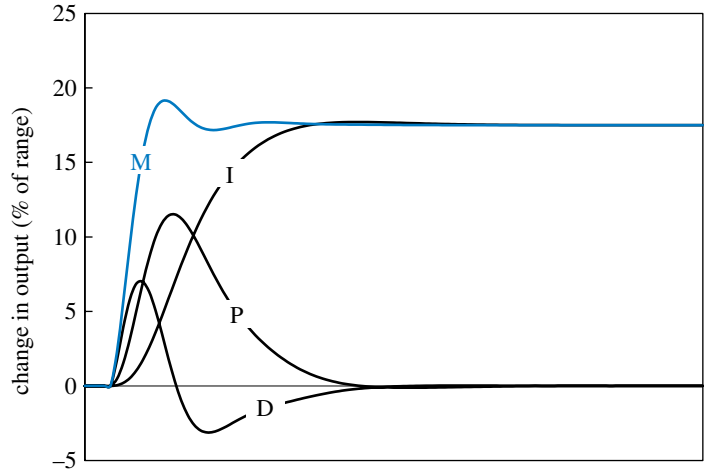

Figure 3.21 *Response to a load change (both algorithms)*

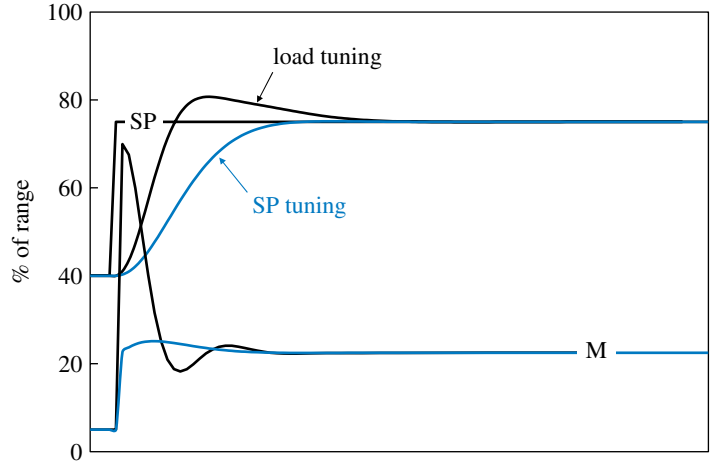

Figure 3.22 *Response to a SP change (P-on-E algorithm)*

We have to consider which algorithm and tuning combination should be used if the controller is the secondary of a cascade. Such controllers are subject to SP changes when the SP of the primary is changed but also when the primary takes corrective action during a load change. Unlike a primary, a secondary controller will be subject to frequent SP changes. Figure 3.17 shows (theoretically) that the proportional-on-error algorithm should be used in the secondary, since this will marginally outperform the proportional-on-PV version. One could make the same argument if MPC is installed, since the controllers it manipulates effectively become secondaries of a cascade.

We will cover later different measures of control performance but the most commonly used is ITAE. The higher the value of ITAE, the poorer the controller is at eliminating the error. Figure 3.23 shows the impact that switching from proportional-on-PV to proportional-on-error has on ITAE. Both algorithms

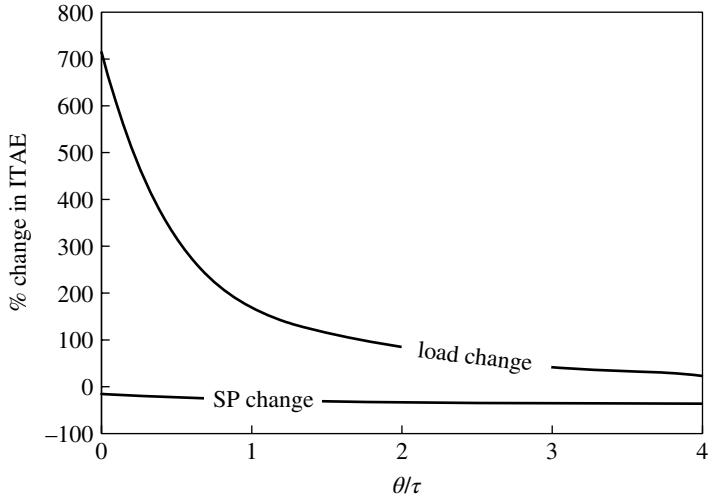

Figure 3.23 *% change in ITAE when switching from P-on-PV to P-on-E algorithm*

have been tuned for SP changes. As expected, use of the proportional-on-error for SP changes would result in an improvement in ITAE. While this might at first appear significant, it should be compared against the much larger degradation when the process is subjected to a load change. Secondaries of cascades generally have a very small θ/τ ratio and so the ITAE during a SP change would be reduced by about 20% whereas that for a load change would increase by around 600%. We must also consider the case when the primary controller is out of service. The secondary will then experience mainly load changes and incur an unnecessarily large ITAE. Further, in the interest of standardisation, universally adopting the proportional-on-PV algorithm (tuned for SP changes) would be advantageous.

Some DCS have the option of automatically switching between the two algorithms as the cascade is switched between auto and manual. While this might seem a good idea, it is advisable to disable this feature. The DCS does not make the necessary change to tuning constants when it switches algorithm. Of course it would possible to add this feature as a customisation but that would involve maintaining two sets of tuning. Given that one set is likely to be rarely used, there is a danger of it not being updated to reflect any process or instrumentation changes. This switching also has an impact when step-testing the secondary SP in order to develop tuning constants for the primary. If the switching is enabled then the proportional-on-PV algorithm will be used during step-testing but the proportional-on-error will be used when the cascade is commissioned. This also applies when step-testing for the later addition of MPC. After the primary (or MPC) is commissioned the dynamics could be quite different from those found in the step-test and may cause performance problems. On systems where this facility cannot be disabled, the proportional-on-error version must be used when the primary (or MPC) is in service and special arrangements have to be made to 'trick' the DCS into using this algorithm during step-testing.

Figure 3.23 does however show that, if the number SP changes and load changes are similar, the advantage of the proportional-on-PV algorithm tends to zero as the θ/τ ratio becomes large. Here we should use the proportional-on-error algorithm if there are more frequent SP changes. However, as we shall see in Chapter 7, there are techniques superior to the PID algorithm for processes with such long deadtimes. While these may not always be applicable, the occasions on which the proportional-on-error algorithm is justified will be very rare.

Some DCS include an algorithm described as the *two degrees of freedom controller*. This has the form

$$\Delta M = K_c\left[\left(x_n - x_{n-1}\right) + \frac{ts}{T_i}E_n + \frac{T_d}{ts}\left(y_n - 2y_{n-1} + y_{n-2}\right)\right] \tag{3.83}$$

where

$$x = PV - \alpha SP \qquad \text{and} \qquad y = PV - \beta SP \tag{3.84}$$

The algorithm can also be described as having *set-point weighting*. The values of α and β can be set by the engineer to a value in the range 0 to 1. We will show in Chapter 15 that their addition is the equivalent of passing SP changes through a lead-lag algorithm (described in Chapter 6). For the proportional action the lead-to-lag ratio is α and, for the derivative action, it is β. Setting both parameters to 1 will give the controller, as described by Equation (3.33), while setting them both to 0 will give the recommended form of the controller as described by Equation (3.82). It is possible to use values between 0 and 1 but there is little benefit in doing so. Optimising the tuning for SP changes will always result in a value for α of 1 (and higher if permitted). However, this ignores the requirement that the controller should handle both SP and load changes well with the same tuning constants. For this to be achievable, α must be 0. One could argue that, since the response to load changes is unaffected by the value chosen for α, the controller could be tuned for load changes and α then adjusted to give the best possible response to a SP change. This is illustrated in Figure 3.24. However, as we will describe in more detail later, this would require us to know the dynamics of the load change. This is rarely practical since process disturbances usually have multiple sources, each with different dynamics. Similarly the value chosen for β has no effect on the response to load changes. Its optimum value will be that which generates the maximum permitted derivative spike; so β will be zero if no spike is permitted.

For systems which do not support the proportional-on-PV algorithm, the proportional-on-error version can be used with the same tuning constants provided that the SP is filtered to ensure that excessive control action is not taken when it is changed. We show in Chapter 15 that for analog control the filter is a simple

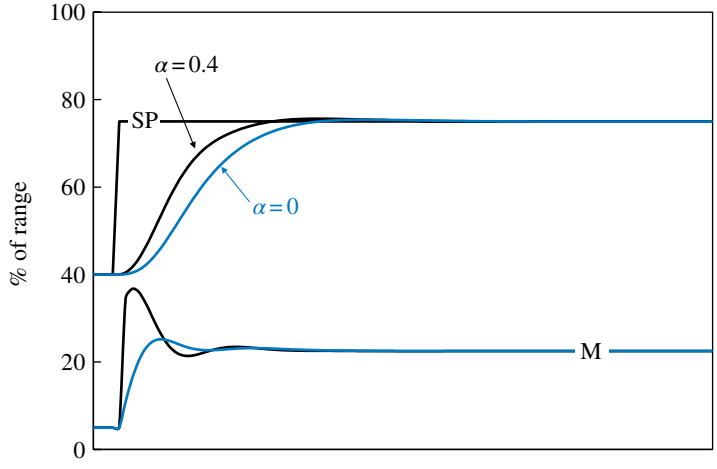

Figure 3.24 *Effect of SP weighting*

lag with its time constant set equal to the integral time (T_i) used in the controller. Indeed, some DCS vendors use the term *SP lag* as the parameter used to configure the proportional-on-PV algorithm. This can also be used for digital control provided the scan interval (ts) is small compared to the integral time. If not, the precise form of the filter depends on how the digital algorithm is defined. As also shown in Chapter 15, for the algorithm described by Equation (3.39), the filtered SP (SP^*) should be derived from the actual SP (SP) as follows:

$$SP_n^* = \left(\frac{T_i}{T_i + ts}\right) SP_{n-1}^* + \left(\frac{ts}{T_i + ts}\right) SP_n \tag{3.85}$$

Like the proportional-on-error algorithm, the proportional-on-PV version will, if set up as proportional-only, result in an offset. However, contrary to Equations (3.13) and (3.20), the offset will be equal to ΔSP because the proportional action no longer responds to SP changes. Equations (3.15) and (3.22) will still apply for load changes.

Another option available in some systems is an *integral-only* algorithm. This has the form

$$\Delta M = \frac{ts}{T_i} E_n \tag{3.86}$$

In the same way that derivative action amplifies noise, integral action attenuates it. The full position discrete form of the algorithm, Equation (3.45), shows how integral action is based on the sum of all previous values of the error. This averages out any noise. This can be seen in Figure 3.25 which shows the effect of adding measurement noise to the PV controlled in Figure 3.19; the integral component of the controller output is virtually noise-free. For SP changes this integral-only algorithm will respond identically to the proportional-on-PV, derivative-on-PV PID algorithm – provided the value used for T_i is that same as K_c/T_i in the PID algorithm. However, in the absence of proportional and derivative actions, the algorithm will respond more slowly to load changes. So, while it was once commonly used in flow controllers, it is difficult to see what advantage it offers.

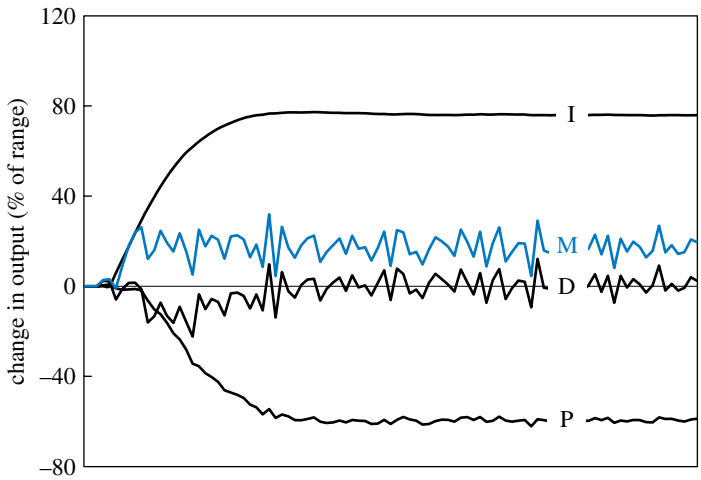

Figure 3.25 *Impact of noise on each part of PID controller*

3.8 Nonstandard Algorithms

Most of the algorithms presented in this chapter can be represented by the general algorithm

$$\Delta M = a_1 PV_n + a_2 PV_{n-1} + a_3 PV_{n-2} + \ldots + b_1 SP_n + b_2 SP_{n-1} + b_3 SP_{n-2} + \ldots \tag{3.87}$$

So, for example, equating coefficients with the basic PID control defined by Equation (3.33) gives

$$a_1 = K_c\left[1 + \frac{ts}{T_i} + \frac{T_d}{ts}\right] \qquad a_2 = -K_c\left[1 + \frac{2T_d}{ts}\right] \qquad a_3 = K_c\frac{T_d}{ts} \tag{3.88}$$

$$b_1 = -K_c\left[1 + \frac{ts}{T_i} + \frac{T_d}{ts}\right] \qquad b_2 = K_c\left[1 + \frac{2T_d}{ts}\right] \qquad b_3 = -K_c\frac{T_d}{ts} \tag{3.89}$$

Similarly, equating coefficients with the preferred PID control defined by Equation (3.82) gives

$$a_1 = K_c\left[1 + \frac{ts}{T_i} + \frac{T_d}{ts}\right] \qquad a_2 = -K_c\left[1 + \frac{2T_d}{ts}\right] \qquad a_3 = K_c\frac{T_d}{ts} \tag{3.90}$$

$$b_1 = -K_c\frac{ts}{T_i} \qquad b_2 = 0 \qquad b_3 = 0 \tag{3.91}$$

However, even in this case, there are more coefficients than conventional tuning constants. So while the coefficients can always be derived from the tuning constants, the reverse is not necessarily true. It is feasible to optimise the coefficients to obtain the best possible controller performance but it is unlikely that it would be possible to convert these to tuning constants to be used in any of the algorithms available in the DCS. If we consider Equation (3.87) a necessary condition for ΔM to be zero, when the process is at steady state with $PV = SP$, is that the sum of the six coefficients must be zero. However, this does not necessarily mean that (for example) a_1 and b_1, as they do in Equations (3.88) and (3.89), must sum to zero. If, when optimised, they did not then the resulting controller cannot be of the PID form and using the PID form would necessarily result in poorer control.

This begs the question of why retain the PID algorithms when the general algorithm can be tuned to outperform any of them. Part of the answer is that there would no longer be an obvious connection between observing that the controller response might be improved and knowing which coefficient(s) to adjust to bring about the improvement. Trial-and-error tuning would become far more time-consuming because of this and the increase in the number of parameters to adjust. Indeed, most engineers would have difficulty achieving even stable control using the equivalent of a PI controller.

While currently restricted to the academic world, there is also the *fractional order* version of the PID algorithm. This is most conveniently described in Laplace form.

$$M = K_c\left[1 + \frac{1}{T_i s^a} + T_d s^b\right]E \tag{3.92}$$

The coefficients a and b, which would be set to 1 in the conventional PID algorithm, are additional tuning constants – typically restricted to the range 0 to 2. Theoretically, the algorithm will outperform the conventional

version in situations where the process cannot be precisely characterised by a single lag (which is theoretically the case for most processes). However, as we will show later, industry already struggles with optimally tuning three parameters. Adding two more will only exacerbate the problem and, even if optimally defined, are unlikely to noticeably improve control performance.

3.9 Tuning

It is probably fair to say that the vast majority of PID controllers in the process industry are not optimally tuned. The majority of tuning is completed using experience and trial-and-error. While this may not adversely affect process performance when the process dynamics are very short, it does become an issue otherwise. It is not the intention of this book that rigorous model identification and tuning be applied to every controller. Improving a fired heater fuel flow controller so that it reacts to a SP change, say, in 5 seconds as opposed to 10 seconds will have a minor impact on the control of the temperature which has dynamics measured in minutes. However, adopting a rigorous approach to the temperature controller is likely to be well worth the effort.

It is unlikely that the control engineer will find a published controller tuning method that will meet the needs of the process. This is despite a considerable amount of research work. In 2000 a survey [5] identified, for self-regulating processes, 81 published methods for tuning PI controllers and 117 for PID controllers. For integrating processes, it also found 22 methods for PI control and 15 for PID control. Every one of these methods has at least one flaw.

The published methods described in this chapter are included primarily to draw the engineer's attention to what limitations might be encountered and permit assessment of any other method offered.

Some tuning methods are based on the *damping ratio*. This is explained in detail in Chapter 15. Damping can describe either open loop or closed loop behaviour. The terminology originates from the analysis of spring/damper combinations as used in vehicle suspension systems. An *overdamped* system, as shown in Figure 3.26, approaches the new steady state gradually without overshooting. An *underdamped* system will exhibit overshoot. A *critically damped* system is one as close as possible to being underdamped without overshoot. An *undamped* system is one which oscillates at constant amplitude.

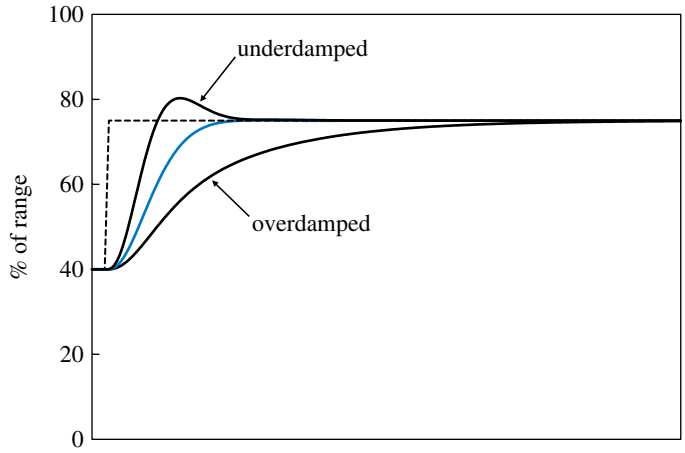

Figure 3.26 *Damping*

An underdamped system has to be at least second order. In the process industry underdamped processes are rare. One example would be the rapid closure of a pressure control valve at the end of a long length of pipework containing liquid travelling quickly. The pressure upstream of the valve would temporarily increase above its steady state value. Known as *fluid hammer*, this can require that pipelines be given special attention. Certain types of reactor can also exhibit open loop underdamped response. However, once a controller is added to an overdamped process, underdamping can be brought about by too tightly tuning the controller.

3.10 Ziegler-Nichols Tuning Method

Perhaps the most well known and most frequently published method is that by Ziegler and Nichols [1]. They developed an open loop method in addition to the more frequently published closed loop method. The criterion used by both methods is the *quarter-decay ratio*. This is illustrated in Figure 3.27; following a disturbance (in this case a SP change), the PV response should be slightly oscillatory with the height of the second peak being one quarter of that of the first. The reciprocal of the decay ratio is known as the *subsidence ratio*.

The nomenclature used by Ziegler and Nichols is somewhat different from what we use today. To avoid confusion we will update it to current terminology. If we address first the closed loop method, the technique involves starting with a proportional-only controller and adjusting its gain until the PV oscillates with a constant amplitude. We record the gain at which this is achieved – known as the *ultimate gain* (K_u) and measure the period of oscillation – known as the *ultimate period* (P_u). This is shown in Figure 3.28; the amplitude of both the PV and the MV is constant and they are in exact anti-phase.

One of the problems in applying the method is the practicality of achieving sustained oscillation. Even if triggered by a very small change in SP, the amplitude can be very large and potentially harmful to the process. The preferred proportional-on-PV controller, if configured as proportional-only, will not respond to SP changes and so the oscillation must be triggered by a load change. This is not always straightforward.

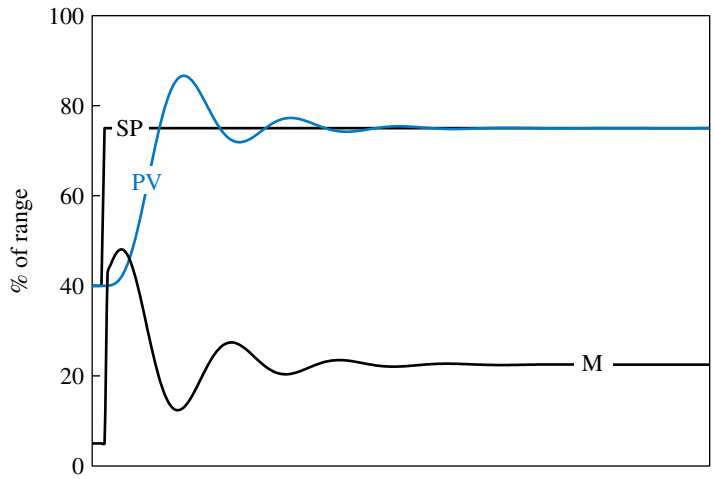

Figure 3.27 *Quarter decay ratio*

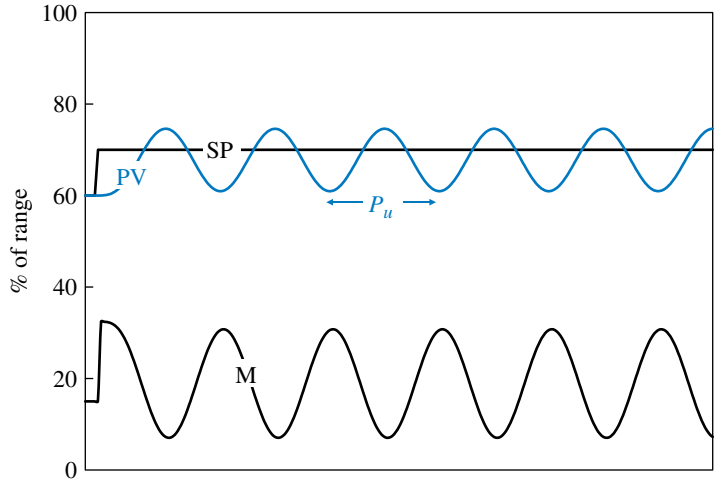

Figure 3.28 *Sustained oscillation*

However, it is possible to predict K_u and P_u from the process dynamics. For a self-regulating process, as derived in Chapter 15

$$K_u = \frac{-1}{K_p \cos\left(\frac{2\pi\theta}{P_u}\right)} \tag{3.93}$$

$$\frac{2\pi\tau}{P_u} + \tan\left(\frac{2\pi\theta}{P_u}\right) = 0 \tag{3.94}$$

Solving Equations (3.93) and (3.94) is not trivial and requires an iterative approach. For K_u to have the same sign as K_p, P_u must be between 2θ and 4θ. A good starting point for the iteration is given by the formula developed by Lopez, Miller, Smith and Murrill [6].

$$K_u = \frac{2.133}{K_p}\left[\frac{\theta}{\tau}\right]^{-0.877} \qquad \text{for} \qquad 0 \le \frac{\theta}{\tau} \le 1 \tag{3.95}$$

The following equation may be used, although its accuracy also degrades somewhat as θ/τ exceeds 1.

$$K_u = \frac{1}{K_p}\frac{1+e^{-\theta/t}}{1-e^{-\theta/\tau}} \tag{3.96}$$

Alternatively P_u can be estimated from Figure 3.29 and the result used in Equation (3.93), or K_u can be determined directly from Figure 3.30.

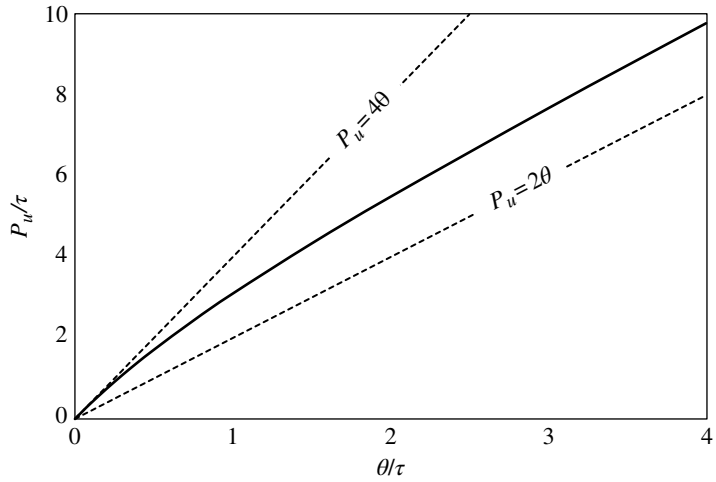

Figure 3.29 *Estimating the ultimate period from process dynamics*

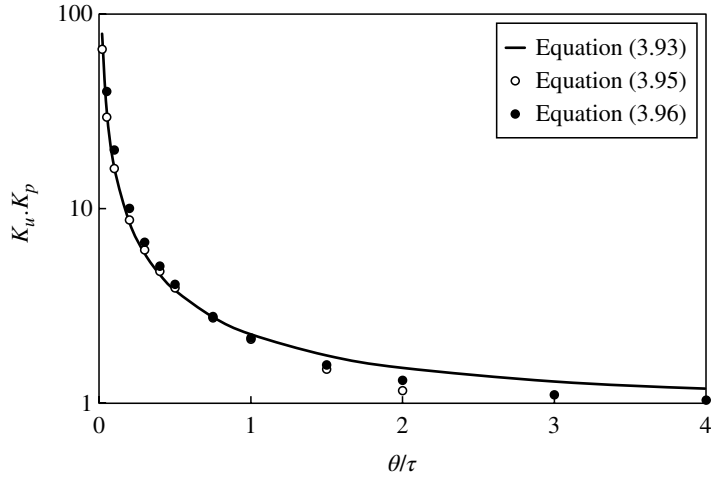

Figure 3.30 *Estimating the ultimate gain from process dynamics*

For an integrating process the solution is considerably easier.

$$K_u = \frac{\pi}{2K_p\theta} \tag{3.97}$$

$$P_u = 4\theta \tag{3.98}$$

An alternative approach to adjusting K_p by trial-and-error to attain sustained oscillation is to apply the *Relay Method* [7]. This involves first selecting acceptable low and high values of the MV (MV_{low} and MV_{high}) that are either side of the current operating point. The MV is switched automatically between these two values when the PV crosses a target set close to the current operation. This will set up a sustained

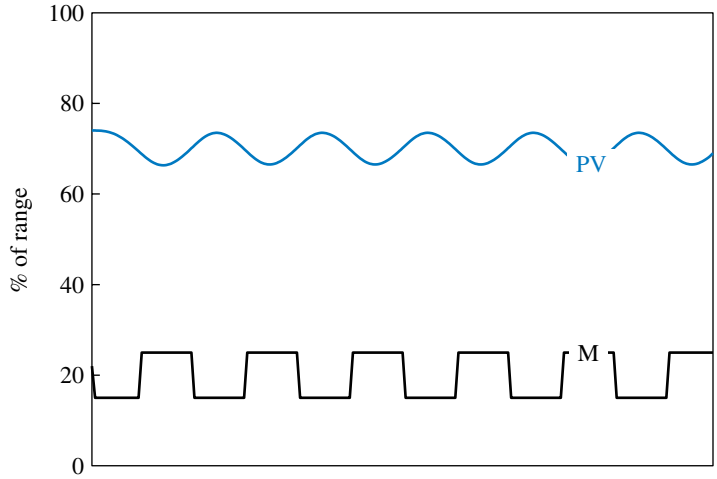

Figure 3.31 *Relay method*

oscillation in the PV. The period of this oscillation will be P_u. The ultimate gain is derived from peak-to-peak variation of the PV.

$$K_u = \frac{4\left(MV_{high} - MV_{low}\right)}{\pi\left(PV_{high} - PV_{low}\right)} \tag{3.99}$$

Figure 3.31 gives an example where the target for the PV is set at 70% and the low and high values for the MV are set at 15% and 25% respectively. The resulting period of oscillation is the same as that in Figure 3.28.

Once K_u and P_u are known, Ziegler and Nichols provide simple calculations for the derivation of tuning constants. For a P-only controller

$$K_c = 0.5K_u \tag{3.100}$$

Because a proportional-only controller will never reach SP, the quarter decay is determined with respect to the steady state condition. The reciprocal of the coefficient, in this case the reciprocal of 0.5, is known as the *gain margin*. It is the factor by which the controller gain can be increased before the controller becomes unstable. A proportional-only controller tuned according to the Ziegler-Nichols method will therefore have a gain margin of 2.

For a PI controller

$$K_c = 0.45K_u \qquad T_i = \frac{P_u}{1.2} \tag{3.101}$$

For a PID controller

$$K_c = 0.6K_u \qquad T_i = \frac{P_u}{2} \qquad T_d = \frac{P_u}{8} \tag{3.102}$$

One might argue that, since the controller is in automatic mode when the closed loop test is performed, any difference in control algorithm is taken into account. This is partially true in that it compensates for the change from analog to digital control. The delay introduced by the scan interval will add to the process deadtime so that the process will begin to oscillate at a controller gain less than required with

analog control – resulting in a lower value for K_c. However, since only proportional action is used to identify K_u, any other changes to the algorithm are not taken into account.

A number of others have suggested slightly different approaches. For example, Harriott [8] proposed adjusting K_c in a proportional-only controller until quarter-decay, rather than sustained oscillation, is achieved. The period of oscillation (P), which will be greater than P_u, is then used in Equation (3.103) to determine T_i and T_d. With these in place, K_c is then readjusted to give quarter-decay.

$$T_i = \frac{P}{1.5} \qquad T_d = \frac{P}{6} \tag{3.103}$$

The quarter-decay tuning criterion is now dated. It is unlikely to be perceived now as a well-tuned controller. It gives substantial overshoot on both the PV and the MV with little advantage to achieving steady state as soon as possible. Some have suggested using different coefficients to achieve a more acceptable response. However, the underlying problem is that K_u and P_u (or P) do not uniquely define the process. For a self-regulating process, they can be derived from K_p, θ and τ but the reverse is not true. There are many combinations of dynamics which would give the same value of K_u and P_u. For the same reason there are many combinations of PID tuning constants that will give the required performance. Ziegler and Nichols dealt with this by fixing the T_d/T_i ratio at 0.25. However, this was not necessary for an integrating process. Because only two dynamic constants are normally used to characterise it, there is a unique relationship as shown by Equations (3.97) and (3.98).

The Ziegler-Nichols open loop tuning technique is an extension of the steepest slope method that we described in Chapter 2 as a means of obtaining the process dynamics. The method, as originally documented, involves calculating the slope (R) of the steepest slope and then dividing this by the step change (ΔMV) to give the *unit reaction rate* (R_1). However, we can update the method to use the actual process dynamics. Without deadtime, a first order self-regulating process is described by

$$\Delta PV = K_p \Delta MV \left(1 - e^{-t/\tau}\right) \tag{3.104}$$

Differentiating

$$\frac{d(\Delta PV)}{dt} = \frac{K_p \Delta MV}{\tau} e^{-t/\tau} \tag{3.105}$$

For a first order process, the rate of change of PV is a maximum when $t = 0$, so

$$R = \frac{K_p \Delta MV}{\tau} \qquad \text{and} \qquad R_1 = \frac{K_p}{\tau} \tag{3.106}$$

Adapting the formulae given by Ziegler and Nichols we get, for a P-only controller,

$$K_c = \frac{\tau}{K_p \theta} \tag{3.107}$$

For a PI controller

$$K_c = \frac{0.9\tau}{K_p \theta} \qquad T_i = \frac{\theta}{0.3} \tag{3.108}$$

And for a PID controller

$$K_c = \frac{1.2\tau}{K_p\theta} \qquad T_i = \frac{\theta}{0.5} \qquad T_d = \frac{\theta}{2} \tag{3.109}$$

For an integrating process, modifying Equation (2.54) gives

$$\frac{d(\Delta PV)}{dt} = K_p \Delta MV \tag{3.110}$$

And so R_1 is the same as K_p and the tuning formulae become, for a P-only controller

$$K_c = \frac{1}{K_p\theta} \tag{3.111}$$

For a PI controller

$$K_c = \frac{0.9}{K_p\theta} \qquad T_i = \frac{\theta}{0.3} \tag{3.112}$$

And for a PID controller

$$K_c = \frac{1.2}{K_p\theta} \qquad T_i = \frac{\theta}{0.5} \qquad T_d = \frac{\theta}{2} \tag{3.113}$$

Given that the ideal version of the PID algorithm was not available until the advent of electronic controllers, Ziegler and Nichols could not have used this to develop their tuning method. While in theory they would have used the interactive version, they actually used a pneumatic controller which approximated to this. In fact, the mathematical definition of the interactive algorithm came after pneumatic controllers were designed. To obtain the tuning for the ideal algorithm we can apply Equations (3.62), (3.63) and (3.64) to Equation (3.102).

$$K_c = 0.75K_u \qquad T_i = \frac{5P_u}{8} \qquad T_d = \frac{P_u}{10} \tag{3.114}$$

Applying them to Equation (3.113) gives

$$K_c = \frac{1.5\tau}{K_p\theta} \qquad T_i = \frac{\theta}{0.4} \qquad T_d = \frac{2\theta}{5} \tag{3.115}$$

For the ideal controller, the T_d/T_i ratio reduces from 0.25 to 0.16, but the 25% increase in K_c will make the controller even more aggressive than the quarter-decay behaviour.

What is often not appreciated is that Ziegler and Nichols developed their tuning method by injecting steps into the controller output. The method is therefore designed for load rather than SP changes. As we saw earlier in this chapter, the resulting tuning will then be too severe for SP changes unless the proportional-on-PV algorithm is used. While the tuning will still likely be too aggressive, it is not as bad as one

Table 3.1 *Chien et al. tuning method*

	PV overshoot = 0%						PV overshoot = 20%					
	SP change			Load change			SP change			Load change		
	K_c	T_i	T_d	K_c	T_i	T_d	K_c	T_i	T_d	K_c	T_i	T_d
P	$\frac{0.3\tau}{K_p\theta}$			$\frac{0.3\tau}{K_p\theta}$			$\frac{0.7\tau}{K_p\theta}$			$\frac{0.7\tau}{K_p\theta}$		
PI	$\frac{0.35\tau}{K_p\theta}$	1.2τ		$\frac{0.6\tau}{K_p\theta}$	4θ		$\frac{0.6\tau}{K_p\theta}$	τ		$\frac{0.7\tau}{K_p\theta}$	2.3θ	
PID	$\frac{0.6\tau}{K_p\theta}$	τ	0.5θ	$\frac{0.95\tau}{K_p\theta}$	2.4θ	0.42θ	$\frac{0.95\tau}{K_p\theta}$	1.4τ	0.47θ	$\frac{1.2\tau}{K_p\theta}$	2θ	0.42θ

would be led to believe by testing it with the wrong algorithm. As with the closed loop method, others have suggested the use of modified calculations. For example, Chien et al. [9] recognised this problem and presented different tuning for SP and load changes applied to the ideal PID algorithm (see Table 3.1).

3.11 Cohen-Coon Tuning Method

Another published method frequently quoted is that by Cohen and Coon [10]. It used the same method as Ziegler and Nichols to obtain the process dynamics, as shown in Figure 2.24. Their tuning technique was developed entirely theoretically assuming a truly first order process with deadtime and the ideal version of the PID controller. Like Ziegler-Nichols tuning was developed for load, rather than SP, changes. To overcome the problem of this being achievable with multiple sets of tuning constants they introduced the *self-regulation index* (μ). They defined this using the same terminology as Ziegler-Nichols. Converting it to current nomenclature, and applying Equation (3.105), shows that μ is the deadtime-to-lag ratio (θ/τ). This is also known as *controllability*. Cohen and Coon published charts from which controller tuning formulae could be derived. The formulae they published were developed from these charts to meet the quarter-decay criterion.

Those familiar with the technique may not immediately recognise the formulae. The self-regulation index has been replaced with the deadtime-to-lag ratio, but at some stage many of the coefficients were converted by another to fractions approximating to the original values. Almost every reproduction of the method presents these fractions. These conceal some of the features of the method so, while not significantly affecting the resulting tuning constants, we have here presented the original formulae.

For P-only control

$$K_c = \frac{1}{K_p}\frac{\tau}{\theta}\left[1.03 + 0.35\frac{\theta}{\tau}\right] \tag{3.116}$$

For PI control

$$K_c = \frac{1}{K_p}\frac{\tau}{\theta}\left[0.9 + 0.083\frac{\theta}{\tau}\right] \qquad T_i = \theta\left[\frac{0.9 + 0.083\frac{\theta}{\tau}}{0.27 + 0.6\frac{\theta}{\tau}}\right] \tag{3.117}$$

For PD control

$$K_c = \frac{1}{K_p}\frac{\tau}{\theta}\left[1.24 + 0.16\frac{\theta}{\tau}\right] \qquad T_d = \theta\left[\frac{0.34 - 0.11\dfrac{\theta}{\tau}}{1.24 + 0.16\dfrac{\theta}{\tau}}\right] \tag{3.118}$$

For PID control

$$K_c = \frac{1}{K_p}\frac{\tau}{\theta}\left[1.35 + 0.25\frac{\theta}{\tau}\right] \qquad T_i = \theta\left[\frac{1.35 + 0.25\dfrac{\theta}{\tau}}{0.54 + 0.33\dfrac{\theta}{\tau}}\right] \qquad T_d = \theta\left[\frac{0.5}{1.35 + 0.25\dfrac{\theta}{\tau}}\right] \tag{3.119}$$

If both are tuned according to this method, we can compare the performance of a PI controller with that of a PID controller. We will define, in the next section, ITAE as a measure of how poorly a controller performs. Figure 3.32 shows that, when the θ/τ ratio is large, this parameter is considerably worse for the PID controller. This is occasionally quoted as evidence (wrongly) that derivative action is less effective for deadtime-dominated processes. This results from the omission made by almost all those who reproduce the Cohen-Coon method; the tuning method is applicable only when the θ/τ ratio is less than 1.

3.12 Tuning Based on Penalty Functions

There are a number of tuning tables published based on four different tuning criteria. In addition to the previously mentioned integral over time of the absolute error (ITAE), there is *integral of the absolute error (IAE)*, *integral of the square of the error (ISE)* and *integral over time of the square of the error (ITSE)*. Each of these is a form of penalty function representing the size and duration of error. The tuning methods

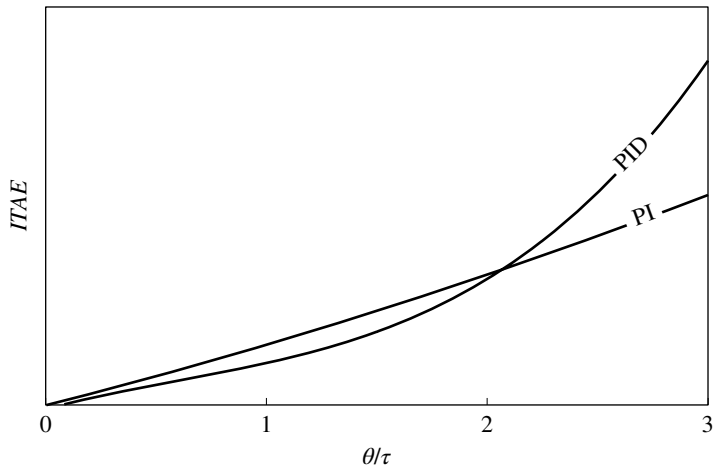

Figure 3.32 *Performance of Cohen–Coon tuning method*

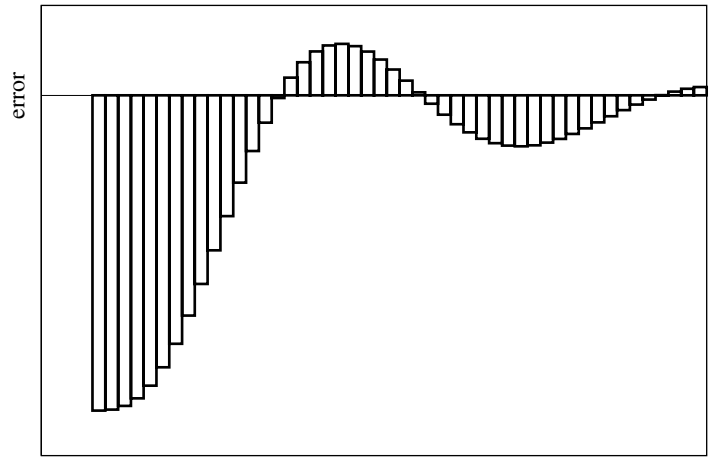

Figure 3.33 *Integral of absolute error*

aim to minimise this penalty. Figure 3.33 shows the IAE. The area between the PV and the SP comprises a series of rectangles of width *ts* (the scan interval) and height |*E*| (the absolute value of the error). The sum of the areas of these rectangles is the IAE. We remove the sign of the error when integrating; otherwise positive errors would be cancelled by negative errors, so that even a sustained oscillation would incur a penalty close to zero. The penalty functions are defined as follows.

$$IAE = \int_0^\infty |E|.dt = ts\sum_{i=1}^{\infty}|E_i| \tag{3.120}$$

$$ISE = \int_0^\infty E^2.dt = ts\sum_{i=1}^{\infty}E_i^2 \tag{3.121}$$

$$ITAE = \int_0^\infty |E|t.dt = ts\sum_{i=1}^{\infty}|E_i|t_i \tag{3.122}$$

$$ITSE = \int_0^\infty E^2t.dt = ts\sum_{i=1}^{\infty}E_i^2t_i \tag{3.123}$$

Minimising *ISE* is equivalent to minimising the variance (σ^2) and thus the standard deviation (σ) of the controller error. But the tuning generated using this penalty function (and also *IAE*) will result in a controller that eliminates as fast as possible the large error, that exists immediately after the deadtime expires, at the expense of causing slightly oscillatory behaviour for some time after the disturbance. The addition of the time since the start of the disturbance (t) in Equations (3.122) and (3.123) provides a weighting factor so that small errors existing a long time after the disturbance make a contribution to the penalty function similar in magnitude to a large error at the start of the disturbance. However, if working dimensionlessly, the absolute value of the error never exceeds 1; so squaring a small error gives a penalty very close to zero. This undermines the advantage of the time weighting in *ITSE*. Figure 3.34 illustrates the difference in controller performance from applying the two time-weighted functions.

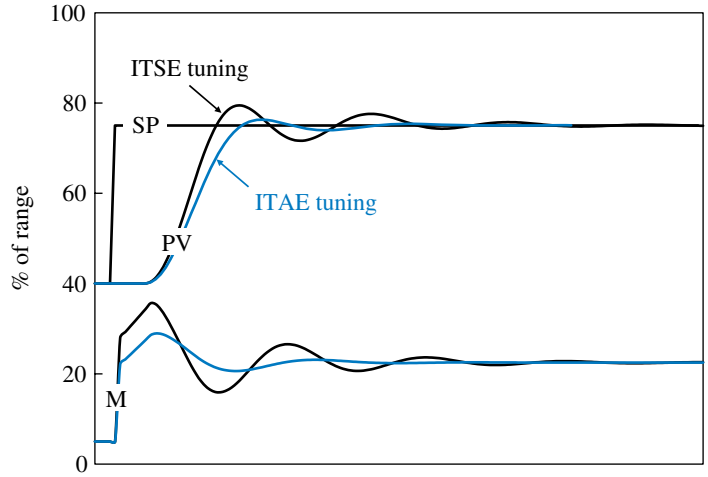

Figure 3.34 *Comparison between ITAE and ITSE tuning*

Tables 3.2 to 3.6 give details of the tuning method developed by Smith, Murrill and others [6,11]. It assumes that the θ/τ ratio is between 0 and 1. Tables 3.2 to 3.4 give tuning designed for load changes. With these it is important to use the proportional-on-PV control algorithm so that the controller does not give an excessive response to SP changes. Tables 3.5 and 3.6 give tuning for SP changes; the method assumes that the proportional-on-error algorithm is used. While it is tempting to explore the use of the method for θ/τ

Table 3.2 *Lopez, Miller, Smith and Murrill (IAE, load change)*

	$K_c = \frac{A\left[\frac{\theta}{\tau}\right]^B}{K_p}$		$T_i = \frac{\tau}{A\left[\frac{\theta}{\tau}\right]^B}$		$T_d = A\left[\frac{\theta}{\tau}\right]^B \tau$	
	A	B	A	B	A	B
P	0.902	–0.985				
PI	0.984	–0.986	0.608	–0.707		
PID	1.435	–0.921	0.878	–0.749	0.482	1.137

Table 3.3 *Lopez, Miller, Smith and Murrill (ISE, load change)*

	$K_c = \frac{A\left[\frac{\theta}{\tau}\right]^B}{K_p}$		$T_i = \frac{\tau}{A\left[\frac{\theta}{\tau}\right]^B}$		$T_d = A\left[\frac{\theta}{\tau}\right]^B \tau$	
	A	B	A	B	A	B
P	1.411	–0.917				
PI	1.305	–0.959	0.492	–0.739		
PID	1.495	–0.945	1.101	–0.771	0.560	1.006

Table 3.4 *Lopez, Miller, Smith and Murrill (ITAE, load change)*

	$K_c = \frac{A\left[\frac{\theta}{\tau}\right]^B}{K_p}$		$T_i = \frac{\tau}{A\left[\frac{\theta}{\tau}\right]^B}$		$T_d = A\left[\frac{\theta}{\tau}\right]^B \tau$	
	A	B	A	B	A	B
P	0.490	–1.084				
PI	0.859	–0.977	0.674	–0.680		
PID	1.357	–0.947	0.842	–0.738	0.381	0.995

Table 3.5 *Smith and Murrill (IAE, SP change)*

	$K_c = \frac{A\left[\frac{\theta}{\tau}\right]^B}{K_p}$		$T_i = \frac{\tau}{A + B\frac{\theta}{\tau}}$		$T_d = A\left[\frac{\theta}{\tau}\right]^B \tau$	
	A	B	A	B	A	B
P						
PI	0.758	–0.861	1.020	–0.323		
PID	1.086	–0.869	0.740	–0.130	0.348	0.914

Table 3.6 *Smith and Murrill (ITAE, SP change)*

	$K_c = \frac{A\left[\frac{\theta}{\tau}\right]^B}{K_p}$		$T_d = A\left[\frac{\theta}{\tau}\right]^B \tau$		$T_i = \frac{\tau}{A + B\frac{\theta}{\tau}}$	
	A	B	A	B	A	B
P						
PI	0.586	–0.916	1.030	–0.165		
PID	0.965	–0.855	0.796	–0.147	0.308	0.929

ratios greater than 1, it can fail dramatically. For example, if θ/τ is greater than 3.2 and Table 3.5 used to tune a PI controller, the resulting integral time will be negative.

While less often reproduced, numerous methods have been published based on a similar approach. The precursor to that presented here, known as the *3C method* [12], gets its name from imposing three constraints on the controller tuning. While the first of these is quarter-decay, to deal with the problem that this can be met by multiple sets of PI tuning constants, it introduced a second constraint of minimising IAE. For full PID the third constraint is that the term $K_pK_cT_d/\tau$ must be 0.5 which, as Equation (3.119) shows, is the same criterion used in the Cohen-Coon method. The formulae are applicable to load changes and values of θ/τ between 0.1 and 1. Outside of this range, much larger values of $K_pK_cT_d/\tau$ are required.

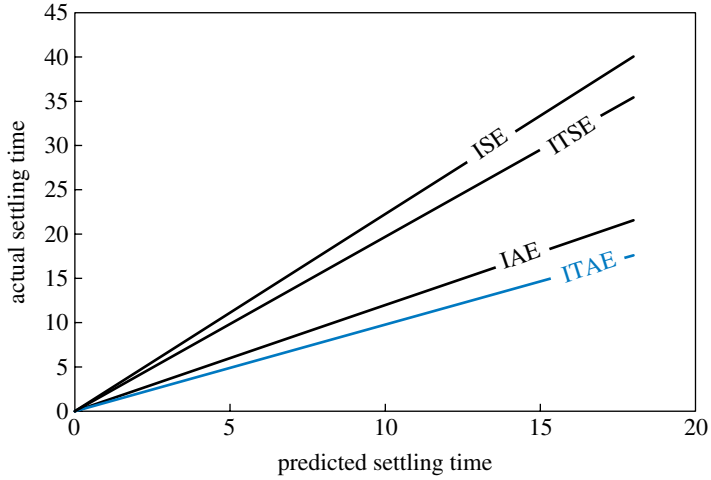

Figure 3.35 *Impact of choice of penalty function on settling time*

One of the most extensive methods [13] provides formulae using each of the same penalty functions, for both SP and load changes, but extended to cover the three main versions of the PID algorithm (interactive, non-interactive and ideal – each with the derivative filter a set at 0.1). As such they are perhaps more applicable to the more commonly installed versions of the algorithm. However, the formulae remain applicable only to values of θ/τ less than 1.

A related penalty function is *settling time*. The 95% settling time is defined as the time taken, following a change in SP, for the error to reduce to 5% of the SP change and then stay within this range. Smuts [14] gives a formula for predicting the minimum achievable value.

$$t_{min} = 2\theta\left(1+\frac{\tau}{\theta+\tau}\right) \tag{3.124}$$

This shows that the minimum settling time will be between 2θ ($\tau = 0$) and 4θ ($\tau \to \infty$). Tests using each of the penalty functions, varying θ and τ each over the range 0–6 minutes, showed that the use of ITAE gave settling times very close to this prediction. As shown in Figure 3.35, the other functions resulted in tuning that gave considerably worse performance. For these reasons ITAE is generally used in this book.

Another term occasionally used is *rise time*. This is the time taken by the PV to first reach the SP. Tuning to minimise rise time will generally result in an oscillatory response and so this criterion is not considered further here.

3.13 Manipulated Variable Overshoot

The main problem with all of the tuning methods presented so far is that they all have the sole objective of reaching the SP as soon as possible. With the exception of some special cases, such as averaging level control described in the Chapter 4, this usually is a requirement. But it is not normally the sole requirement. Figure 3.36 shows (in black) the performance of a controller tuned to meet this aim. However, depending on the process, this might result in excessive adjustments to the MV. In our fired heater example it is unlikely

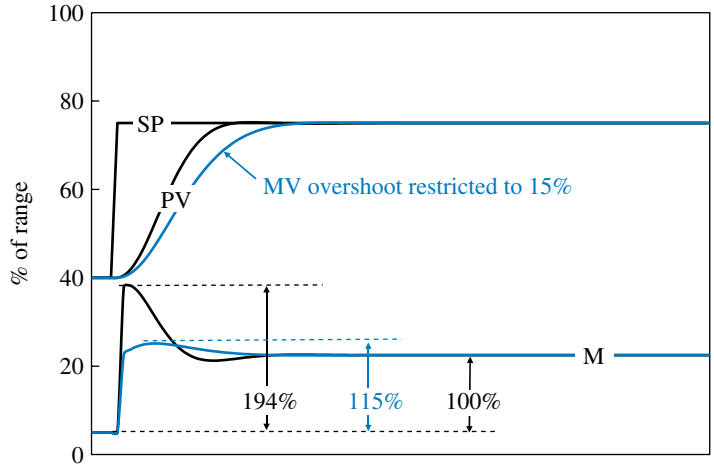

Figure 3.36 *Limiting MV overshoot*

that the fuel could be changed as rapidly as shown without causing a problem on the combustion side. For example it might not be possible to get sufficient air into the firebox as quickly as the rapid increase in fuel would demand. Not doing so would result in incomplete combustion and the controller increasing the fuel even further in an attempt to arrest the resulting fall in temperature. However, on another process, the MV might be some minor flow routed to a section of the process that can tolerate very rapid fluctuations.

If we wish to take MV movement into account we must first define some measure of this. Here we use *MV overshoot* and define this as the percentage by which the MV exceeds the steady state change required to meet the new SP. In our example the MV must move by 17% of range (from 5% to 22%) in order for the PV to achieve its SP. If not restricted, the MV temporarily reaches a maximum value of 38% – giving a maximum change of 33%. In this case overshoot would be calculated as 100(33/17 – 1) or 94%. If we were to restrict this to 15% then the maximum change in the MV would be 17% multiplied by 1.15. Added to the starting value of 5% gives a maximum value of around 25%. Figure 3.36 shows (as the coloured line) the effect on the response of the PV if we apply this limit.

It is important we distinguish between the MV overshoot and the PV overshoot. A number of published tuning methods permit definition of the allowable PV overshoot. However, this does not satisfy the need to place a defined limit on the movement of the MV. An easy check to determine whether a tuning method takes account of this is to determine what tuning constants would be derived if θ is set to zero. Each of the methods above would give the result

$$K_c \to \infty \qquad T_i = 0 \qquad T_d = 0 \tag{3.125}$$

In effect each method suggests that controller gain be set to maximum, the integral action be set to maximum (remembering it uses the reciprocal of T_i) and the derivative action switched off. We might have anticipated the last of these since we have shown that derivative action is only beneficial to SP changes if there is deadtime. The values for K_c and T_i are theoretically correct for analog control. If a process truly has no deadtime (and similarly no scan delay) then increasing controller gain will not cause oscillation. In fact, the tuning recommended would ensure the PV follows the SP immediately. For digital control this can be demonstrated by deriving from Equation (2.16) the model for a first order process with no deadtime.

$$\Delta PV_n = P\Delta PV_{n-1} + (1-P)K_p\Delta MV_{n-1} \qquad \text{where} \qquad P = e^{-ts/\tau} \tag{3.126}$$

If the PV immediately follows the change in SP then, at the next scan

$$\Delta SP = P\Delta PV_n + (1-P)K_p\Delta MV_n \tag{3.127}$$

If we start with the process at steady state then ΔPV_n will be zero. Rearranging gives a proportional-only controller (based on error) with K_c set to $1/(1-P)K_p$.

$$\Delta MV_n = \frac{\Delta SP}{(1-P)K_p} \tag{3.128}$$

The required steady change in MV is given by

$$\Delta MV = \frac{\Delta SP}{K_p} \tag{3.129}$$

The expected % MV overshoot can therefore be determined as

$$\frac{\Delta MV_n - \Delta MV}{\Delta MV} = \frac{100P}{1-P} = \frac{100e^{-ts/\tau}}{1-e^{-ts/\tau}} \tag{3.130}$$

As an example, consider a process with a lag of 1 minute and a controller scan interval of 1 second – giving a MV overshoot of almost 6000%. If we were to restrict the overshoot to 15%, Equation (3.130) tells us that the process lag must be less than half the scan interval – in our example less than 0.5 seconds. In the process industry such dynamics are very unlikely to be encountered.

Figure 3.37 shows just how severe this problem can be (note that the vertical scale is logarithmic). If we were to tune, with no restriction on MV overshoot, any process with a θ/τ ratio less than 1.8 then the overshoot would exceed our nominal 15% limit. This would therefore apply to the vast majority of controllers.

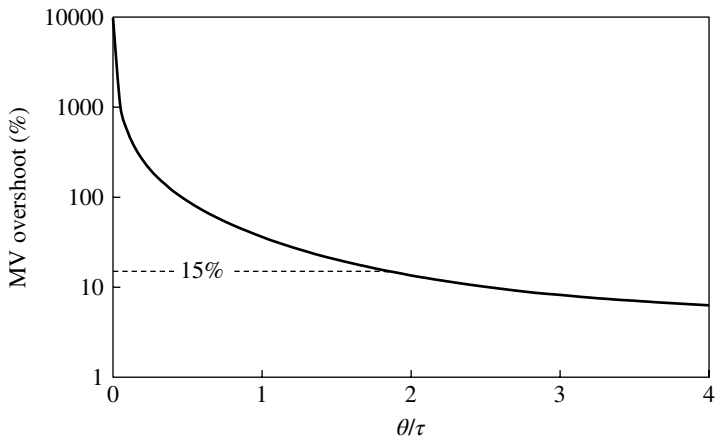

Figure 3.37 *Unrestricted MV overshoot*

If one of the penalty functions described above is used to optimise tuning for a proportional-only controller on a self-regulating process then some restriction on MV overshoot is essential. Minimising the penalty is best achieved by maximising controller gain to minimise offset, even if this causes oscillatory behaviour. Without any restriction the 'optimised' gain will be close to the ultimate gain (K_u). One might consider basing the penalty not on deviation from SP but on how much the PV deviates from its final steady state value. However, such a penalty would be minimised (to zero) by setting the process gain to zero.

On an integrating process an offset will occur, when applying proportional-only control, only if it is subjected to a load change. A limit on MV overshoot will again be essential. However, such a limit cannot be applied to SP changes, no matter what form of controller is used. The steady state change in the MV will be zero and so overshoot cannot be defined as a percentage of ΔMV.

3.14 Lambda Tuning Method

None of the methods so far described give the engineer any way of explicitly limiting movement of the MV. The only approach would be to start with the calculated tuning and adjust by trial-and-error. One of the approaches to address this issue is the Lambda method first introduced by Dahlin [15]. This includes an additional tuning parameter (λ). This is the desired time constant of the process response to a SP change and gives the engineer the facility to make the controller more or less aggressive.

The tuning constants are calculated for a self-regulating process from formulae developed by Chien [16] as

$$K_c = \frac{\tau}{K_p\left(\lambda + \theta\right)} \qquad T_i = \tau \qquad T_d = 0 \tag{3.131}$$

And for an integrating process as

$$K_c = \frac{2\lambda + \theta}{K_p\left(\lambda + \theta\right)^2} \qquad T_i = 2\lambda + \theta \qquad T_d = 0 \tag{3.132}$$

Figure 3.38 shows (as the coloured line) the open loop response for a step change in MV. The closed loop responses of three controllers (with λ set at 0.5τ, τ and 2τ) are shown in black. With λ set to the process lag the closed loop response closely follows the open loop response. If λ is increased then the controller will adjust the MV more slowly, reduce the MV overshoot and increase the time taken for the PV to reach its SP. If λ is decreased below τ, the PV will overshoot the SP.

3.15 IMC Tuning Method

Lambda tuning is an example of *internal model control (IMC)* tuning. It is developed using a technique known as *direct synthesis* [17]. This can be applied to higher order processes and to all types of error-based controllers. Some examples are presented in detail in Chapter 15, but the principle is to synthesise a controller that will respond to a SP change according to a defined trajectory. However, the result may not have the form of the PID algorithm and so approximations have to be made. For example, higher order

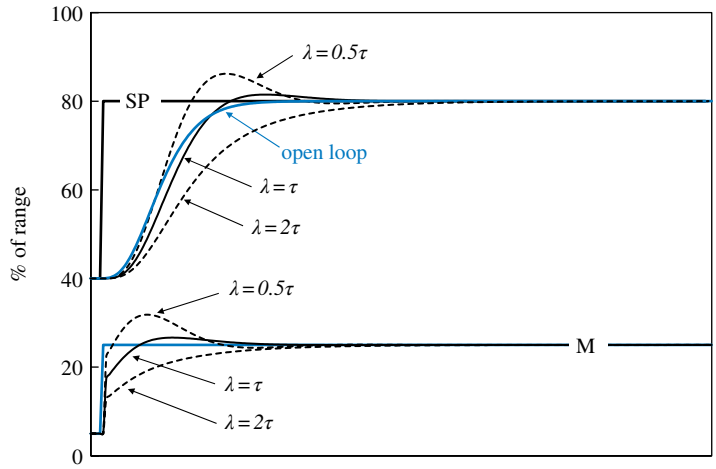

Figure 3.38 *Effect of λ on controller response to SP change*

Table 3.7 *IMC tuning formulae*

	Self-regulating			Integrating		
	K_c	T_i	T_d	K_c	T_i	T_d
PID (ideal)	$\frac{1}{K_p}\frac{\tau+\frac{\theta}{2}}{\lambda+\theta}$	$\tau+\frac{\theta}{2}$	$\frac{\tau\theta}{2\tau+\theta}$	$\frac{1}{K_p}\frac{2\lambda+\theta}{\left(\lambda+\frac{\theta}{2}\right)^2}$	$2\lambda+\theta$	$\frac{\lambda\theta+\frac{\theta^2}{4}}{2\lambda+\theta}$
PID (interactive)	$\frac{1}{K_p}\frac{\tau}{\lambda+\theta}$	τ	$\frac{\theta}{2}$	$\frac{1}{K_p}\frac{2\lambda+\frac{\theta}{2}}{\left(\lambda+\frac{\theta}{2}\right)^2}$	$2\lambda+\frac{\theta}{2}$	$\frac{\theta}{2}$
PI	$\frac{1}{K_p}\frac{\tau}{\lambda+\theta}$	τ		$\frac{1}{K_p}\frac{2\lambda+\theta}{(\lambda+\theta)^2}$	$2\lambda+\theta$	

terms are neglected. Different developers delete these at different stages in the derivation. Further, if the process has deadtime, this would require the controller to take action before the SP is changed. Again different approximations are made to deal with this. This results in slight differences between the resulting tuning formulae. Table 3.7 lists the formula commonly required, but many versions will be found in the literature.

It is possible to use the technique to quantify other parameters in the control algorithm. For example Rice and Cooper [18] developed formula for a – the term used in derivative filtering so that, on those DCS which permit the engineer to change this value, it can also be optimised.

A problem with the method is that it hides the benefit of including derivative action. If a PI and a PID controller are both tuned using the same value of λ then not surprisingly they will follow the same trajectory

in response to a SP change. Indeed, the only difference between the two arises because slightly different approximations are made in developing the tuning formulae. Derivative action already has a much undeserved bad reputation with some engineers without well-publicised methods providing 'evidence' that it has little impact.

While adjusting λ changes the aggressiveness of the controller, its value has to be determined by trial-and-error. While there are several published techniques for selecting λ, there is no predictable relationship between its value and MV overshoot. Under a different set of process dynamics the relationship between λ and MV overshoot will change. This is illustrated in Figure 3.39. The curves were plotted by testing the tuning given in Table 3.7 for the ideal proportional-on-error PID controller on a self-regulating process. They show that the value of λ required to give a required MV overshoot (say 15%) varies as the θ/τ ratio varies.

Of course it is possible from this chart to construct another allowing the engineer to choose a value for λ to give the required MV overshoot. Indeed, this has been done and the result shown as Figure 3.40. And, from this chart, simple formulae could be developed. For example for a 15% MV overshoot.

$$\lambda = 0.88\tau + 0.31\theta \tag{3.133}$$

The same approach can be applied to the interactive proportional-on-error PID controller (with the derivative filter a set to 0.1) also on a self-regulating process. For a 15% MV overshoot the formula becomes

$$\lambda = 0.88\tau - 0.13\theta \tag{3.134}$$

The reader might be concerned that, if θ/τ is greater than 6.77, this equation suggests that λ should be negative. This merely indicates that, with such dynamics, the MV overshoot cannot exceed 15%.

There is, however, no guarantee that the graphs developed, as in Figure 3.40, will be linear. Indeed, extending the technique to a proportional-on-error PI controller on a self-regulating process requires a quadratic relationship to accurately determine the value of λ required for a 15% overshoot.

$$\lambda = 0.83\tau + 0.09\theta + 0.075\frac{\theta^2}{\tau} \tag{3.135}$$

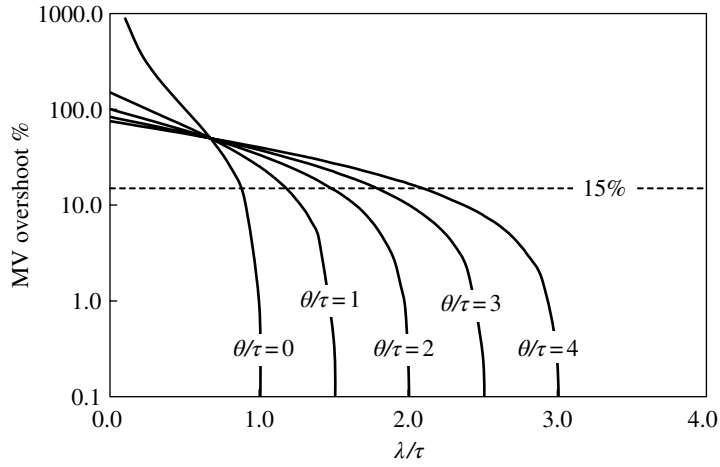

Figure 3.39 *Effect of λ on MV overshoot*

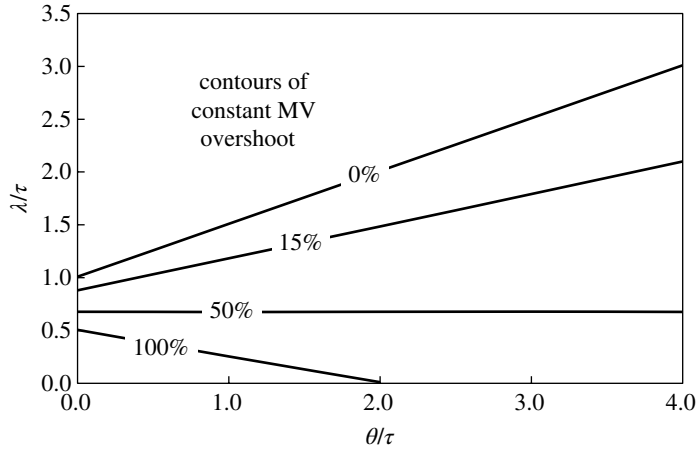

Figure 3.40 *Value of λ necessary to give required MV overshoot*

So, for this approach to be adopted, such charts and formulae would have to be developed for every version of the PID algorithm, for every variation of the tuning formulae, for both integrating and self-regulating processes and for both SP and load changes. The number of charts required would be impractically large. Further the formulae in Table 3.7, for self-regulating processes, show that only K_c changes as λ is changed. We will show later that, to retain optimum tuning, T_i and T_d should also be changed. Because direct synthesis assumes analog control, IMC tuning will fail if applied to processes with dynamics of the same order as the controller scan interval. But the most serious limitation of the method is that, to date, no one has published tuning formulae for the preferred proportional-on-PV, derivative-on-PV algorithm.

3.16 Choice of Tuning Method

All the tuning methods we have covered assume analog control. This is reasonable if the process dynamics are long compared to the scan interval. However, there will be occasions where this is not the case. Compressors, for example, can show vary fast dynamics particularly with surge avoidance controls. Indeed, compressor manufacturers will often specify the use of analog or fast-scanning digital controllers. However, we will show later that it is possible to tune DCS-based controllers for this application, provided the tuning method takes account of the controller scan interval. Similarly discontinuous instruments, such as on-stream analysers, generate measurements relatively infrequently. If a controller using this measurement only takes action when there is a new value, then the scan interval will again be large compared to the process dynamics.

So, to summarise, we should be looking for the following features in a tuning method:

- It is designed for the control algorithm. Our preferred algorithm is the ideal, proportional-on-PV, derivative-on-PV version. The method must also be suited to any DCS-specific features in the algorithm, particularly if these cannot be disabled. For instance the derivative filter term (a) is not adjustable in many DCS and should therefore be taken account of by the tuning technique.
- It permits the engineer to explicitly limit the MV overshoot when required. Methods that do not take this into account will suggest very aggressive tuning as the θ/τ ratio falls. By checking what tuning results

for a zero deadtime process we can determine whether MV movement is taken into account. The engineer-defined performance criterion must be a consistent measure of MV movement, no matter what the process dynamics.

- It is designed for digital rather than analog control. If the scan interval is used in the tuning calculations then this is likely to be the case. But this only becomes an issue if the process dynamics are very fast and approach the scan interval.

It is unlikely that a formula-based approach to tuning is practical. The different versions of the PID controller are not mutually exclusive. We have described three options for the error term included in the integral action. There are the ideal and interactive versions. Derivative action can be based on error or PV. Proportional action can be based on error or PV. This already gives 24 possible combinations without all the DCS-specific enhancements. We have two types of process, self-regulating and integrating, and we have not yet considered higher-order processes such as those with inverse response. We have to accommodate derivative filtering, the choice of tuning criteria (ITAE, etc., MV overshoot), the controller scan interval and whether we want tuning for SP or load changes. We would thus need several hundred sets of tuning formulae to cover all the options. It is not surprising therefore that no published technique has yet proved effective for all situations.

Instead we should learn from the engineers who have largely adopted trial-and-error as their preferred technique. But instead of the time-consuming exercise of performing this on the real process, we can simulate it. This is provided we have, from plant testing, an understanding of the process dynamics. We could, of course, use one of the many commercial computer-based tuning products available. However, care should be taken in selecting one of these. Few would meet the criteria listed above. Alternatively we can develop a simulation of our process and the controller either in code or in a spreadsheet, define the tuning criteria and have the computer optimise tuning to meet these criteria. It is this approach that has been used to generate the following figures.

Of course these figures can lead the engineer into the same trap. They again appear to offer a set of standard approaches to cover all situations. However, the approaches embody all the recommendations developed in this chapter and, where practical, they leave the engineer some flexibility in their application. Used with care, almost any controller can be optimally tuned.

3.17 Suggested Tuning Method for Self-Regulating Processes

Figures 3.41, 3.42 and 3.43 give the recommended tuning for the preferred algorithm (ideal, proportional-on-PV, integral-on-E_n, derivative-on-PV with no derivative filtering). It is assumed that the scan interval is small compared to the process dynamics. The tuning is designed to minimise ITAE subject to a maximum MV overshoot of 15% on a self-regulating process. The points plotted are the results of the trial-and-error simulation technique previously described. Curve fitting gives the following tuning formulae.

$$K_c = \frac{1}{K_p}\left[\left(1.038\frac{\theta}{\tau} + 0.353\right)^{-1.644} + 0.583\right] \tag{3.136}$$

$$T_i = \tau\left[\left(0.588\frac{\theta}{\tau} + 4.164\right)^{0.929} - 2.971\right] \tag{3.137}$$

$$T_d = \tau\left[\left(1.190\frac{\theta}{\tau} + 3.850\right)^{0.487} - 1.857\right] \tag{3.138}$$

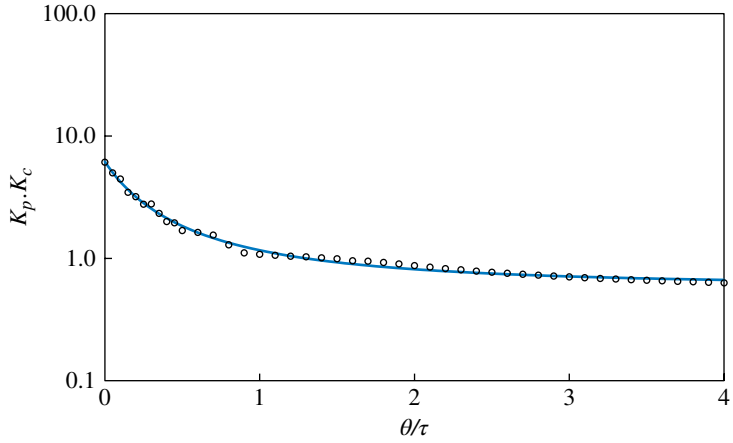

Figure 3.41 *Controller gain*

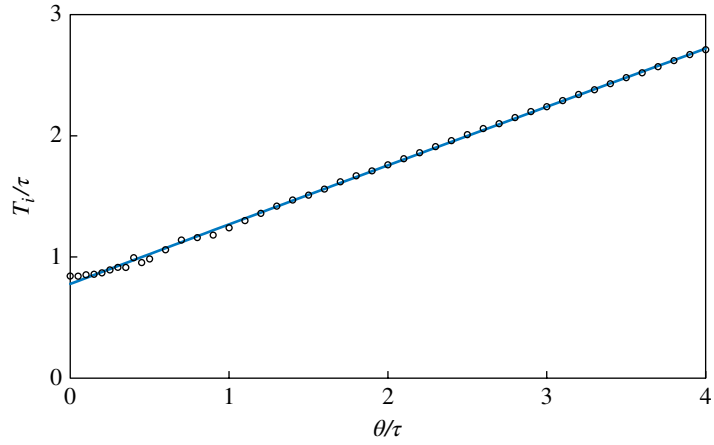

Figure 3.42 *Integral time*

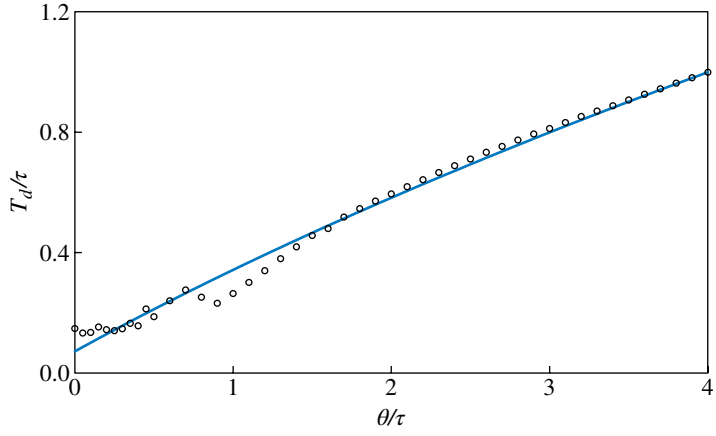

Figure 3.43 *Derivative time*

Figures 3.44, 3.45 and 3.46 show the effect of switching to the more commonly used proportional-on-error algorithm. Using the same approach, the following formulae were developed.

$$K_c = \frac{1}{K_p}\left[\left(5.428\frac{\theta}{\tau}+1.311\right)^{-0.528}+0.267\right] \tag{3.139}$$

$$T_i = \tau\left[\left(0.705\frac{\theta}{\tau}+4.254\right)^{0.848}-2.639\right] \tag{3.140}$$

Derivative action should be included if θ/τ is greater than 0.5.

$$T_d = \tau\left[\left(4.798\frac{\theta}{\tau}+5.107\right)^{0.314}-1.882\right] \tag{3.141}$$

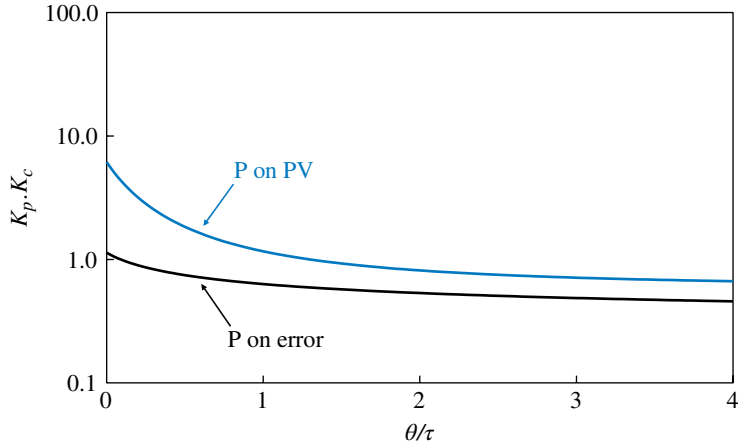

Figure 3.44 *Comparison of algorithms (controller gain)*

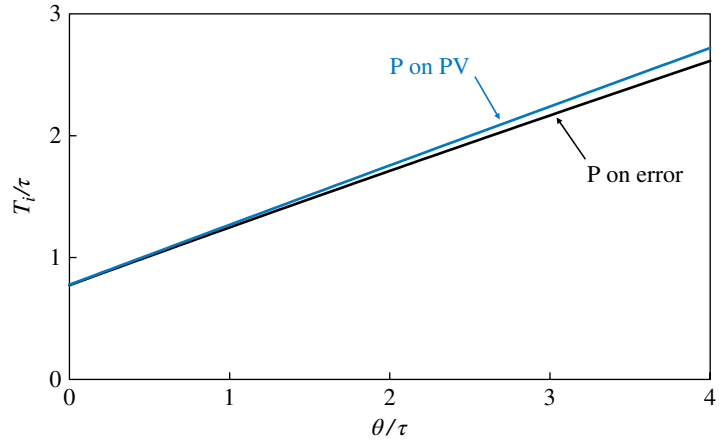

Figure 3.45 *Comparison of algorithms (integral time)*

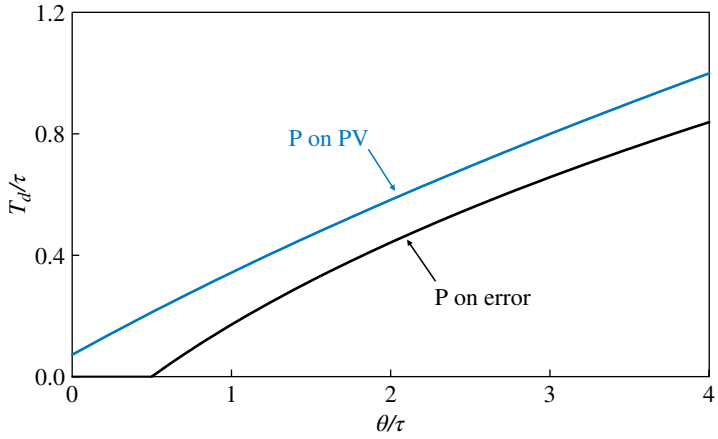

Figure 3.46 *Comparison of algorithms (derivative time)*

This again shows the benefit of selecting the proportional-on-PV algorithm. Equation (3.136) generates a much higher controller gain than Equation (3.139) to give the equivalent response to SP changes. Use of the algorithm will therefore much more rapidly deal with load changes.

A less well-known chart-based tuning method, developed by Marlin and Ciancone [19], is based on similar tuning criteria. It uses the ratio $\theta/(\theta + \tau)$ rather than θ/τ. As a result the dimensionless integral and derivative tuning constants will be generated as $T_i/(\theta + \tau)$ rather than T_i/τ, and $T_d/(\theta + \tau)$ rather than T_d/τ. To compare the results of such a method conversion can be performed as follows:

$$\frac{T_i}{\tau} = \frac{T_i}{\theta+\tau} \times \left(\frac{\theta}{\tau} + 1\right) \quad \text{and} \quad \frac{T_d}{\tau} = \frac{T_d}{\theta+\tau} \times \left(\frac{\theta}{\tau} + 1\right) \tag{3.142}$$

The conditions, under which Equations (3.136) to (3.141) were developed, may not apply to every situation. The following examples show the impact on the charts of:

- changing to tuning for load changes
- impact of ramping SP changes
- relaxing the MV overshoot constraint
- using only PI control
- changing the scan interval

3.18 Tuning for Load Changes

Figures 3.47, 3.48 and 3.49 show the difference between tuning for load and SP changes. We have seen that tuning for load changes can be faster than that for SP changes because the error changes more slowly. The problem with deriving tuning by simulation is that we have to make an assumption about the process dynamics of the PV with respect to the source of the process disturbance. In the absence of any better information, we assume that they are the same as those with respect to the MV. Indeed this is how, by

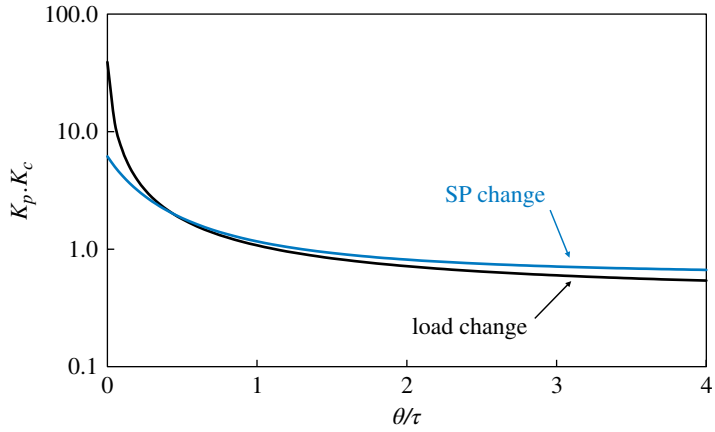

Figure 3.47 *SP versus load tuning (controller gain)*

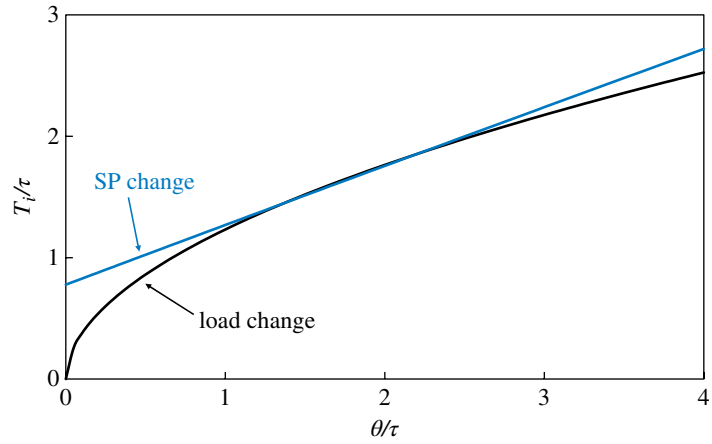

Figure 3.48 *SP versus load tuning (integral time)*

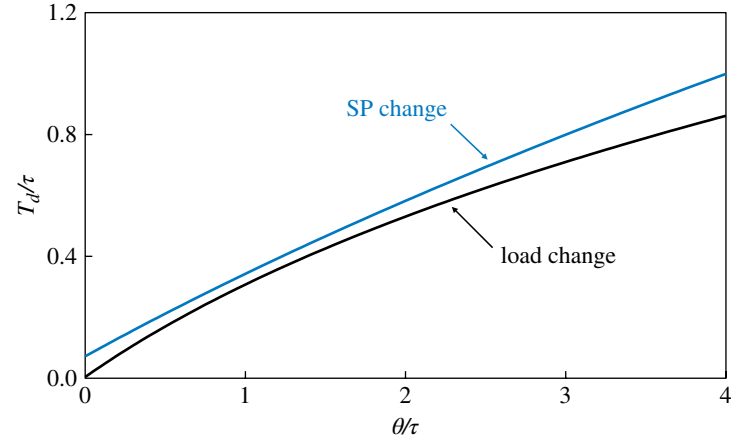

Figure 3.49 *SP versus load tuning (derivative time)*

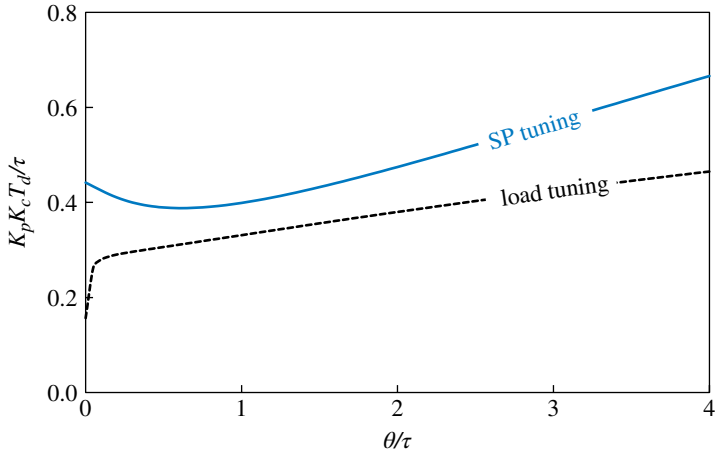

Figure 3.50 *Use of $K_pK_cT_d/\tau$ ratio*

injecting steps into the controller output, the Ziegler-Nichols method was developed. In our example heater, this is the same as saying that the dynamic relationship between outlet temperature and feed rate (or indeed any other load variable) is the same as that between the outlet temperature and fuel rate. This is unlikely to be the case and, almost certainly, different sources of process disturbance are likely to have different dynamics. So any load tuning method must be used with caution. Figure 3.47 suggests that, for processes with a θ/τ ratio of less than about 0.4, we can tune the controller to be more aggressive with load changes. However, it will then over-react to SP changes. For larger θ/τ ratios the tuning for both disturbances is similar. Figure 3.48 shows that we should take the same approach to integral action remembering that, as we increase T_i, integral action is reduced. While Figure 3.49 suggests the same argument does not apply to derivative action, there is only a small difference in T_d for the two cases. Overall then, we should always tune for SP changes.

As described earlier in this chapter, some load tuning methods (such as Cohen-Coon) suggest that the ratio $K_pK_cT_d/\tau$ should be fixed at 0.5. Figure 3.50 shows the optimum value of this factor arising from the tuning charts. It shows that the Cohen–Coon method will always generate tuning that is too aggressive. For SP tuning the ratio is closer to 0.5 but its variation with process dynamics means that the use of a fixed ratio would generally result in non-optimum control performance.

3.19 Tuning for SP Ramps

On many processes, the SP may be changed as ramp rather than a step change. In practice this is performed by making a series of very small steps. One might argue therefore that the proportional-on-error is preferable to the proportional-on-PV algorithm. Figure 3.51 shows a cycle on a batch process where the temperature is held constant for 5 minutes, ramped up 20% of its range over 10 minutes, held constant for 35 minutes, ramped down over 20 minutes and then held constant for 20 minutes. Both algorithms have been tuned using Figures 3.41 to 3.43. As expected the proportional-on-error algorithm performs slightly better. Whether this should be chosen depends on several issues. In this example it tracks the SP over a minute sooner on both of the ramps. The whole cycle lasts 100 minutes, so one

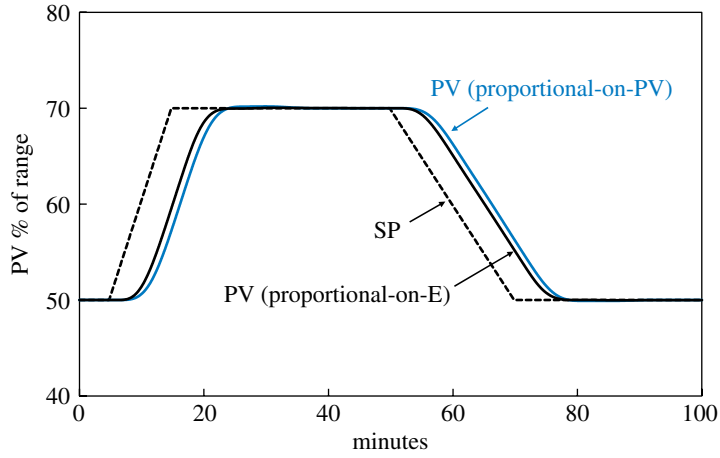

Figure 3.51 *Comparison of algorithms for a ramped SP*

might argue that this could be reduced by over two minutes. If the cycle is a bottleneck on the process then this might permit a production increase of more than 2%.

However, we know that the proportional-on-error algorithm, because of the need to use a smaller controller gain, will respond less well to process disturbances. If these are significant then it may be more profitable to have tighter control of temperature and accept a slightly longer cycle. One should also consider whether the benefits of both might be achieved, for example, by starting the SP ramp a minute sooner or by ramping more quickly.

This also begs the question whether a controller, tuned for step changes in SP, is optimally tuned for a ramped SP. This is similar to the argument for tuning for load changes since these also cause the controller error to accumulate over time. Figure 3.52 shows the proportional-on-PV algorithm optimally tuned for the first ramp. The permitted MV overshoot was not relaxed but nevertheless a substantial increase was

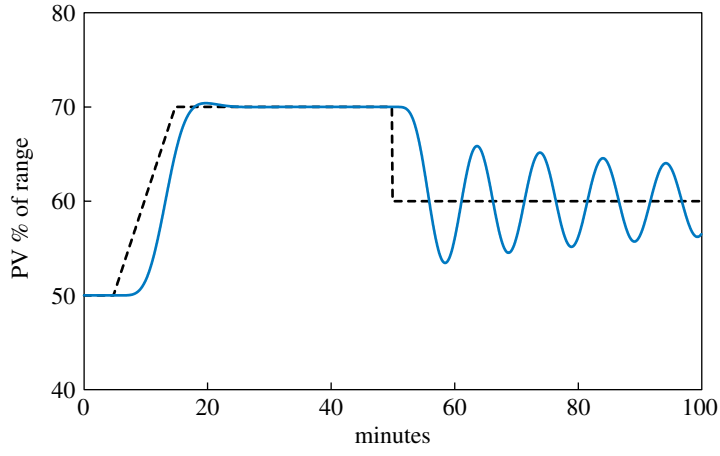

Figure 3.52 *Impact of tuning for ramped SP*

possible in both the proportional and integral actions. As a result the ramp was completed around five minutes sooner. The problem of course is that such tuning would be too fast for steeper ramps. As illustrated in this case the controller would be almost unstable for a step change in SP. It would be similarly oscillatory if load changes cause the error to change more rapidly than the change caused by the SP ramp. Such an approach is thus probably inadvisable. Indeed, if the process can withstand the PV changing more quickly than the SP, then perhaps instead the ramp should be made steeper.

3.20 Tuning for Unconstrained MV Overshoot

Figures 3.53, 3.54 and 3.55 give the tuning for the same controller but this time showing the effect of removing the constraint on MV overshoot. The result is that, as the θ/τ ratio approaches zero, the tuning is the same as that given by many of the published methods, as shown in Equation (3.125).

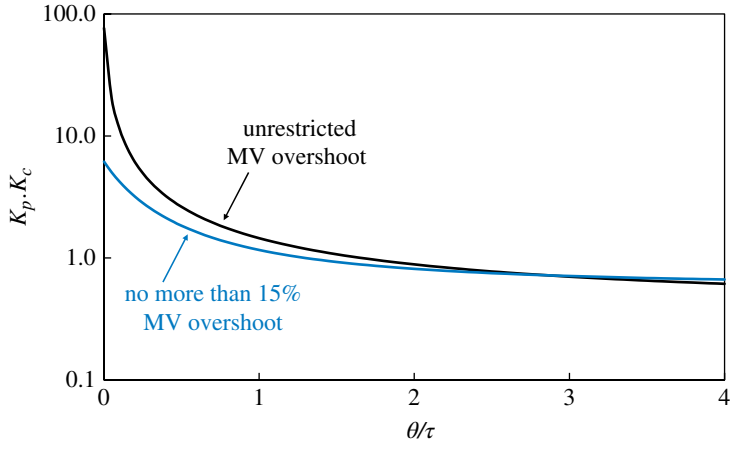

Figure 3.53 *Effect of MV overshoot constraint on controller gain*

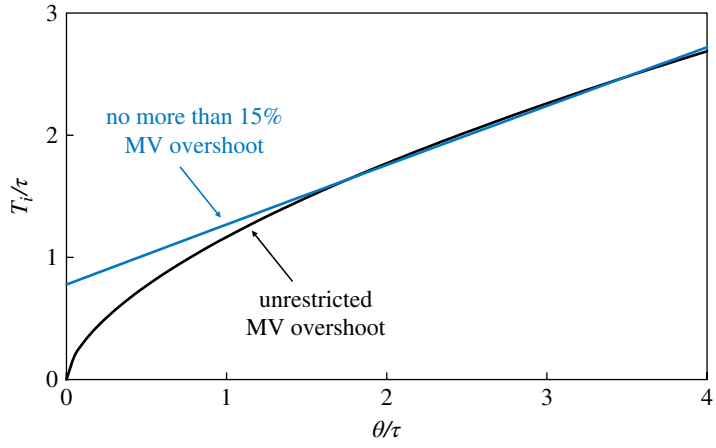

Figure 3.54 *Effect of MV overshoot constraint on integral time*

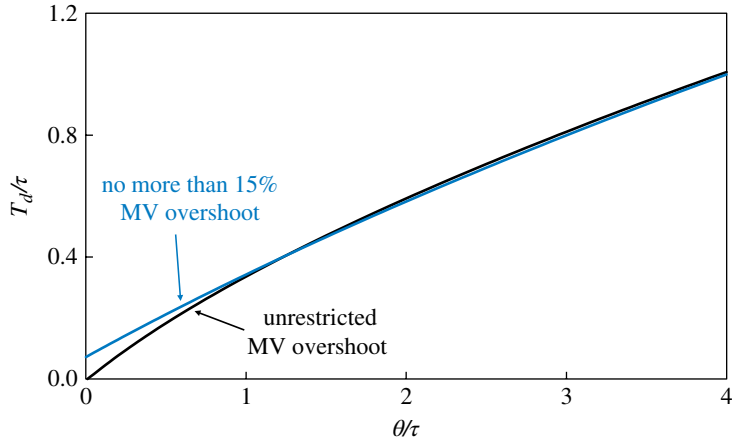

Figure 3.55 *Effect of MV overshoot constraint on derivative time*

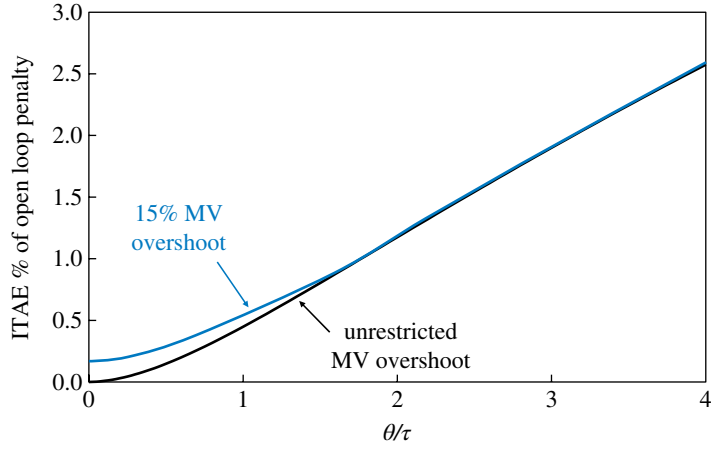

Figure 3.56 *Effect of MV overshoot constraint on ITAE*

While constraining MV movement will clearly slow down the response to disturbances, the effect can be very small. Figure 3.56 shows the impact on ITAE. The chart has been scaled so that 100% ITAE corresponds to the open loop response, i.e. no corrective action is taken over the period in which the process would normally reach steady state ($\theta+5\tau$). Even when θ/τ is close to zero, the impact on ITAE is small. This is partly due to the inclusion of derivative action. Figure 3.55 again demonstrates that this reduces MV overshoot.

3.21 PI Tuning Compared to PID Tuning

Examination of all the methods described in this chapter shows that different tuning is required for the proportional and integral actions if derivative action is added to an optimally tuned PI controller. Figures 3.57 and 3.58 show the effect of switching from PID to PI. They give the tuning for a SP change

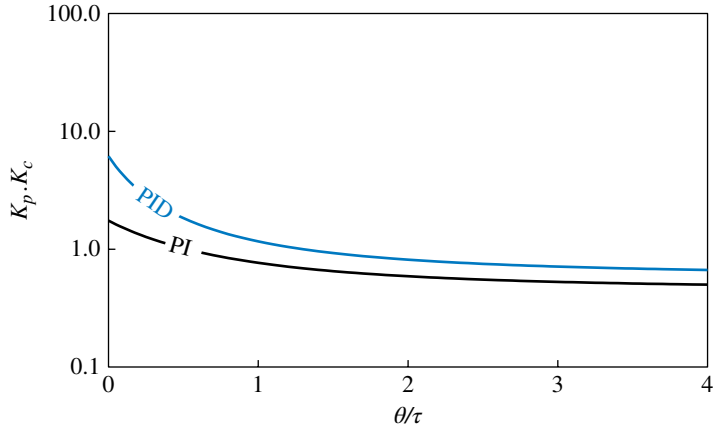

Figure 3.57 *PI versus PID tuning (controller gain)*

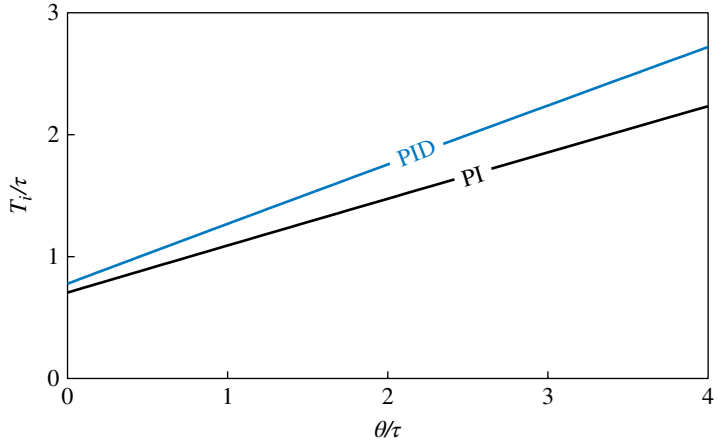

Figure 3.58 *PI versus PID tuning (integral time)*

using the preferred algorithm (proportional-on-PV, integral on E_n). Since there is no derivative action they apply to both the interactive and ideal versions and are not affected by derivative filtering. The tuning is designed to minimise ITAE subject to a maximum MV overshoot of 15% on a self-regulating process.

Figure 3.59 shows what percentage increase in controller gain results from the addition of derivative action. Most published tuning methods suggest that derivative action offers no advantage if the deadtime is zero. However, one of the main purposes of derivative action is to prevent excessive overshoot, so restricting the MV overshoot, e.g. to 15%, dramatically changes its importance. Its inclusion permits the largest increase in controller gain as deadtime approaches zero. As we have seen, the same methods suggest an unacceptably large controller gain and should be rejected in any case.

All tuning methods show that derivative action should be increased as θ/τ increases. Some texts can be somewhat confusing on this issue. They state that derivative action becomes less effective as θ/τ increases and so reach the contradictory (incorrect) conclusion that derivative action should be reduced. As Figure 3.59 shows, derivative action permits a smaller increase in controller gain but the increase is still substantial.

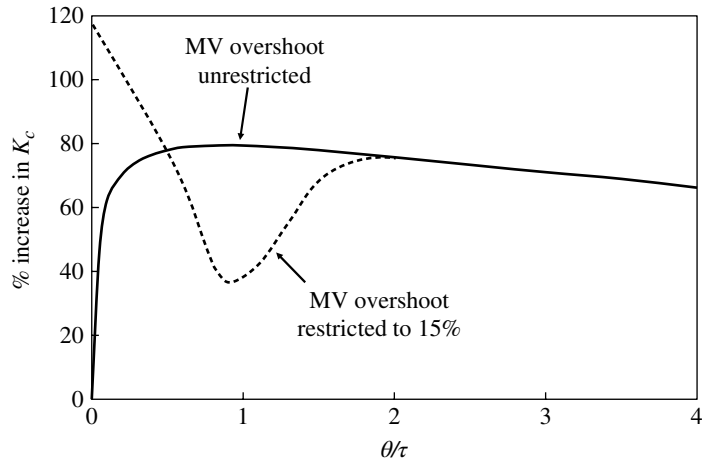

Figure 3.59 *Increase in controller gain from using PID rather than PI control*

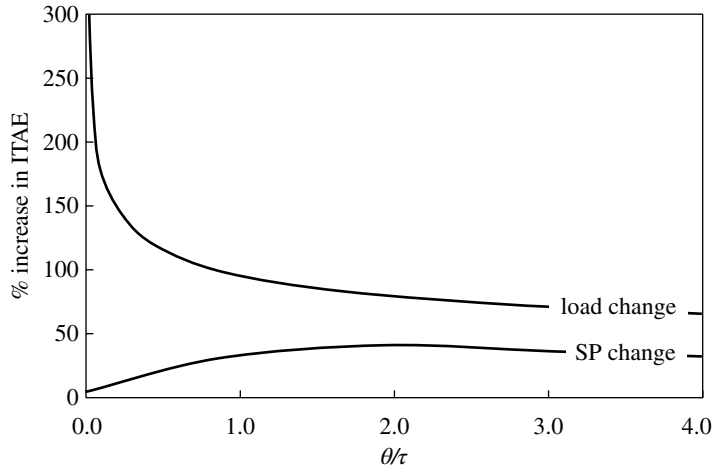

Figure 3.60 *Effect of not using derivative action*

Figure 3.60 shows the worsening of control performance (measured as ITAE) when switching from well-tuned PID to well-tuned PI control. For low θ/τ ratios, the impact on SP changes is relatively minor. However, the adverse effect on load disturbances is substantial, particularly if θ/τ is less than 0.5. Figures 3.57 and 3.58 should therefore only be used to derive PI tuning where derivative cannot be included (e.g. because of spikes caused by step changes in the PV).

3.22 Tuning for Large Scan Interval

In most cases, the scan interval (ts) will be small compared to the process lag (τ) and the controller can be treated as analog. But Figures 3.61 to 3.63 show that the controller can be tuned successfully even when the scan interval approaches the lag. As a modification to some analog controller tuning methods, the developer

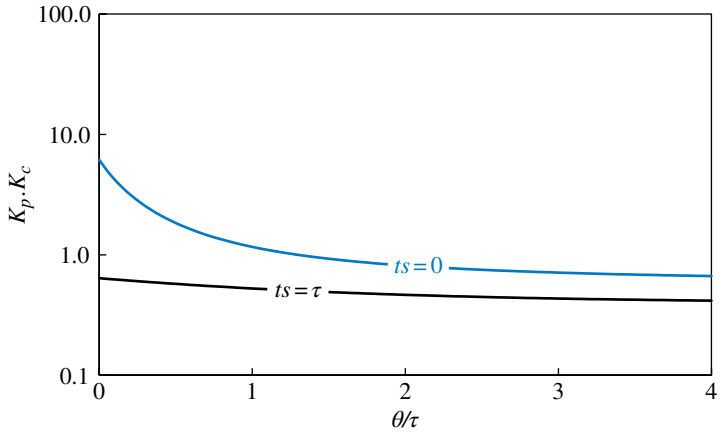

Figure 3.61 *Effect of scan interval on controller gain*

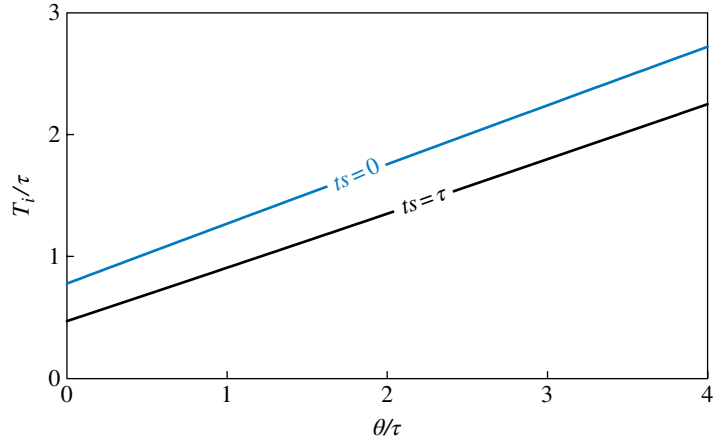

Figure 3.62 *Effect of scan interval on integral time*

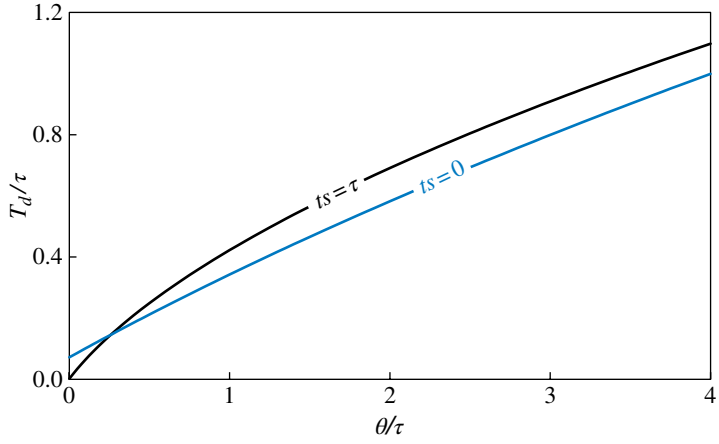

Figure 3.63 *Effect of scan interval on derivative time*

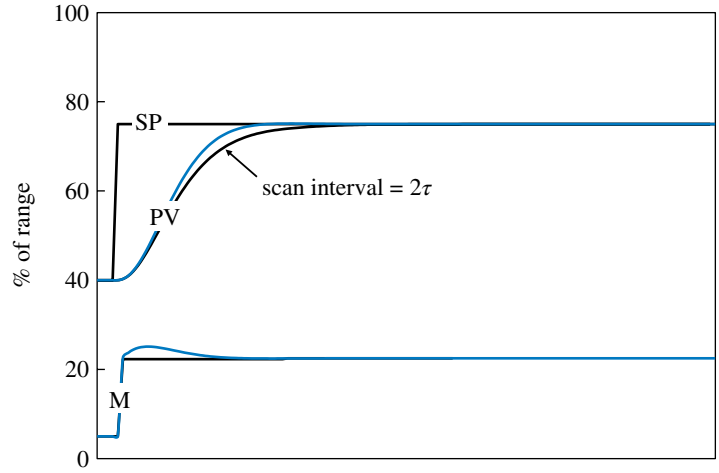

Figure 3.64 *Impact of large scan interval*

has suggested replacing θ in the tuning formulae with $\theta + ts/2$. This is on the basis that digital control will, on average, increase the deadtime by half the scan interval. However, if this were a good approximation, we would expect the curves in Figure 3.61 to be horizontally spaced by a distance of 0.5. Actually the spacing is much larger than this value, showing that the estimate of controller gain is very sensitive to scan interval.

Figure 3.64 shows the impact of increasing the scan interval from zero (analog control) to a value equal to double the process lag (2τ) and retuning the controller to take account of the change. The delay to the PV reaching SP would probably not be noticed on a real process, although perhaps the larger, less frequent, steps now made by the MV would be. What is important is that the controller still behaves well. It is a misconception that increasing scan interval substantially affects controllability.

We are not suggesting here that scan intervals can be dramatically increased, e.g. to alleviate the processing load on a DCS. It is still important that the delay in detecting a SP or load change is not excessive. While we could readily control a process with a lag of (say) 5 minutes, using a controller with a scan interval of 5 minutes, we would not want the controller to take no action for those 5 minutes if a disturbance happens to occur immediately after a scan. What we are suggesting is that a controller scanning, say, every second or two is capable of controlling processes with dynamics of the same order of magnitude.

Scan interval also has an impact on the noise propagated through the controller – ultimately to the control valve. Increasing the scan interval of controller with a noisy PV will, in a given time, mean the controller will respond to fewer noise spikes. Secondly, a larger scan interval means, on average, that rate of change of PV will be lower and so there will be less amplification of the noise by the derivative action. Finally, increasing the scan interval requires the controller gain to be reduced and so noise amplification is further lessened. Li suggests [20] modifying penalty functions, such as ITAE, to take account of valve movement.

$$ITAE' = a\int_0^\infty |E|t.dt + (1-a)\int_0^\infty |\Delta M|dt \tag{3.143}$$

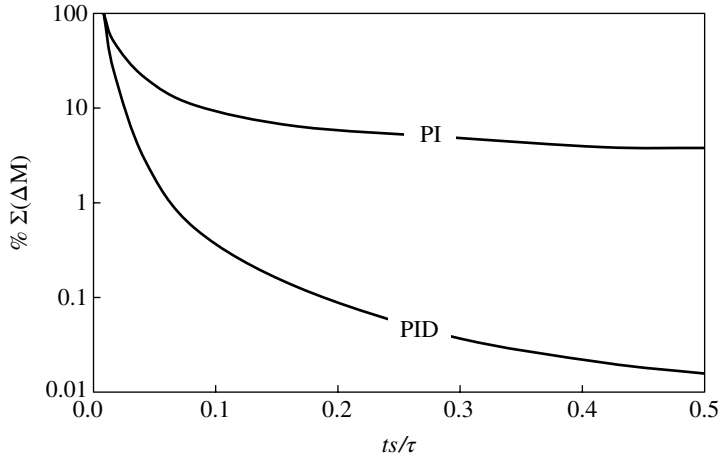

Figure 3.65 *Effect of controller scan interval on valve travel*

The weighting factor (a) is set by the engineer to set the relative importance of, in this case, ITAE versus total valve travel. This is probably too sophisticated an approach. We have seen that increasing scan interval has only a small effect on ITAE. However it can have a dramatic effect, as shown in Figure 3.65, on total valve travel. This was derived from a process with a θ/τ ratio of 0.5 plus noise with a standard deviation of 0.5% of range. Changing these values will cause the curve to move but the underlying shape will be the same. The curve starts at a ts/τ ratio of 1/120 – equivalent to a controller with a scan interval of 1 second on a process with a lag of 2 minutes. Defining the total valve travel under these conditions as 100%, we can see that, for a PID controller, doubling the scan interval reduces it by a factor of 4. Increasing it to 10 seconds reduces valve travel by a factor of over 150.

The impact of scan interval on valve travel is less if no derivative action is used. With a PI controller on the same process it is reduced by half as the scan interval is doubled. Increasing it tenfold reduces it by factor of 10. However, the benefit of even these smaller improvements far outweighs the small increase in ITAE.

On some systems the engineer may have limited flexibility to modify the scan interval. Some systems do not permit changes at all, in others it is modifiable for the whole system but not for individual controllers. However, when the option is available, increasing the scan interval is a very effective way of reducing the transmission of noise to the final control element and should be considered before PV filtering is applied. Conversely upgrading to a newer control system, with a shorter scan interval, will amplify noise and can mislead the engineer into believing there is a problem with the new system.

3.23 Suggested Tuning Method for Integrating Processes

We could include a similar set of charts for tuning controllers on integrating processes. However, since they are all straight line relationships we can more easily represent them as formulae. They assume that K_p has units of min^{-1}, θ has units of minutes and ts has units of seconds. K_c will then be dimensionless; T_i and T_d will have the units of minutes.

While Table 3.8 gives tuning for virtually all possible situations the preferred choice of algorithm and disturbance is listed last.

Table 3.8 *Tuning for integrating process*

Algorithm	Type of algorithm	Proportional based on	Derivative based on	Disturbance	$K_c = \frac{1}{K_p}\left[\frac{60A}{\theta + B\frac{ts}{60}}\right]$		$T_i = A\left[\theta + B\frac{ts}{60}\right]$		$T_d = A\left[\theta + B\frac{ts}{60}\right]$	
					A	B	A	B	A	B
P				load	0.01059	0.552				
P		E		SP	0.00873	0.697				
PI				load	0.00768	0.724	15.5	0.675		
PI		E		SP	0.01185	0.854	4.53	0.578		
PI		PV		SP	0.01208	0.846	3.38	0.543		
PID	interactive			load	0.00988	1.729	6.56	0.696	0.796	1.443
PID	interactive	E	E	SP	0.01503	1.243	2.23	0.867	0.463	0.366
PID	interactive	E	PV	SP	0.01518	1.392	1.96	0.660	0.516	0.468
PID	interactive	PV	E	SP	0.01580	1.288	2.06	0.794	0.509	0.339
PID	interactive	PV	PV	SP	0.01620	1.307	1.95	0.784	0.501	0.347
PID	ideal			load	0.01470	3.422	7.55	0.761	0.351	5.180
PID	ideal	E	E	SP	0.01903	1.496	2.51	0.832	0.374	0.274
PID	ideal	E	PV	SP	0.01947	1.523	2.53	0.784	0.365	0.267
PID	ideal	PV	E	SP	0.02530	2.078	1.34	0.946	0.305	1.153
PID	ideal	PV	PV	SP	0.02153	1.587	2.31	0.785	0.387	0.317

3.24 Measure of Robustness

As discussed in Chapter 2, it is important that the controller design is robust – in that it can accommodate variations in process dynamics without becoming oscillatory. Several tuning packages generate robustness charts. The mathematics involved in their development are presented in Chapter 15. Figure 3.66 shows one such chart for a self-regulating process. The vertical axis shows by what factor the process gain (K_p) may increase before a sustained oscillation occurs. Above this point the controller will be unstable. Horizontally the chart shows by what factor the θ/τ ratio may increase. If it is thought that the process dynamics may vary outside of the stable region then controller tuning should be designed based on a larger value of K_p or θ/τ (or both). Although, preferably, the cause of the changing dynamics should explored and some form of adaptive control implemented – as described later in this chapter and also in Chapter 5.

Some published tuning methods specifically allow for robustness. For example that developed by Marlin and Ciancone [19] assumes that K_p, θ and τ can each vary by 25%. So, if the process dynamics are exactly as measured, the controller performance will not be optimum but it will perform well over the expected range of dynamics.

The curves in Figure 3.67 were derived assuming the controller was tuned according to the method described by Equations (3.136), (3.137) and (3.138). These impose a limit of 15% on MV overshoot. This limit was removed for the tuning used to develop Figure 3.66. This demonstrates a further advantage of limiting MV overshoot. For very low θ/τ ratios the tuning is far more robust. For example, if the controller is designed for θ/τ of 0.2, it will remain stable if θ/τ increases to 0.6. Tuning based on unconstrained MV overshoot will only be stable up a value of 0.3. Similarly, if θ/τ is 1, K_p can increase by a factor of 2.2 if MV overshoot is restricted. Without the restriction the factor reduces to 1.6. As we saw previously, only on processes with θ/τ less than 1.8, does imposing a 15% limit on MV overshoot have an impact on tuning. For this reason the robustness of the controller tuned for θ/τ of 2 is unaffected by relaxing this constraint.

Figure 3.68 shows the stability envelope for an integrating process with a PI controller. The two points show the tuning recommended for SP changes – developed, assuming analog control, from the formulae given in Table 3.8. This tuning recommended is very robust – requiring large increases in either K_p or θ (or both) for the controller to become unstable.

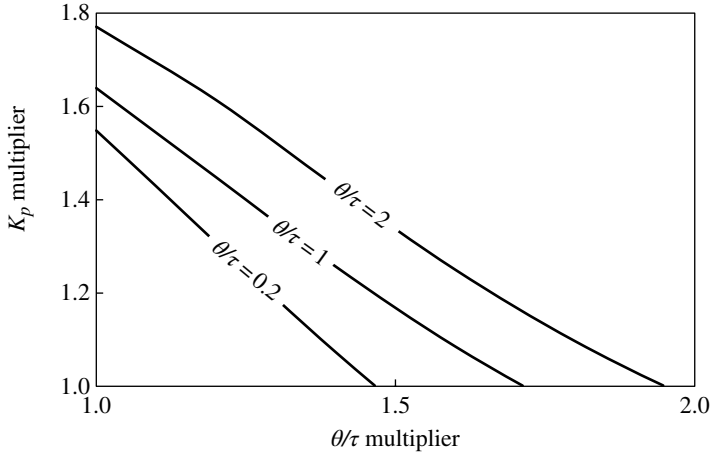

Figure 3.66 *Robustness of a PID controller tuned with no MV overshoot limit*

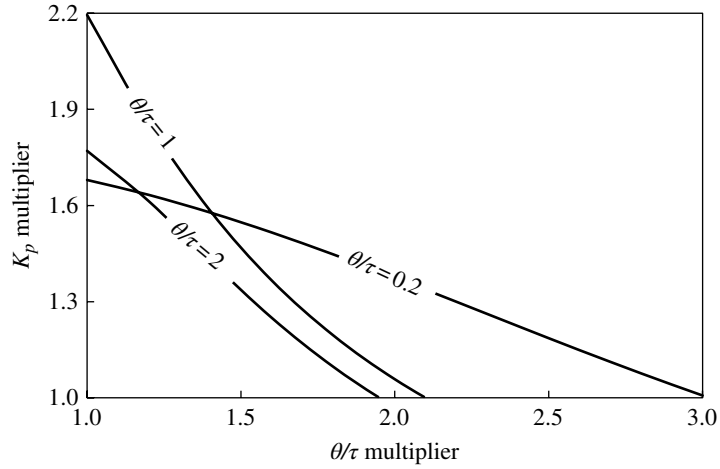

Figure 3.67 *Robustness of a PID controller tuned with 15% MV overshoot limit*

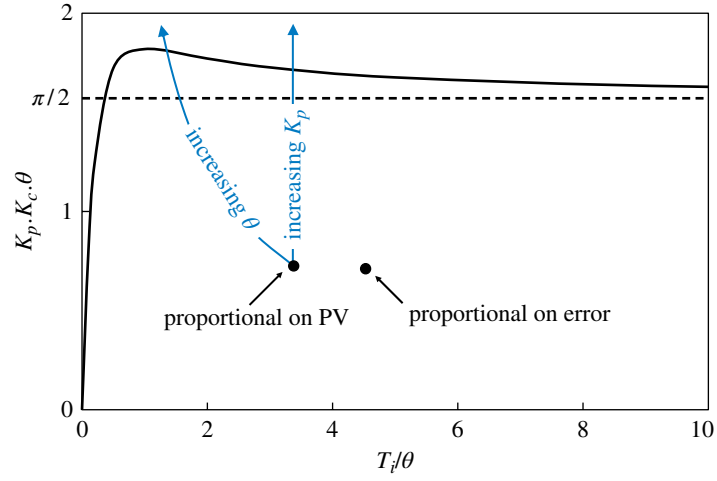

Figure 3.68 *Robustness of a PI controller on an integrating process*

3.25 Implementation of Tuning

Caution should be exercised in implementing new tuning constants. If the tuning has been calculated for an existing controller then comparison with the current tuning might show large changes. If the algorithm has been changed from the proportional-on-error to the proportional-on-PV type then a large increase in K_c is to be expected. It is not unusual for this to increase by a factor of two or three. A change larger than this should be implemented stepwise, testing with intermediate values before moving to those calculated. Similarly, if an existing tightly tuned level controller is re-engineered as an averaging controller, as we shall see in Chapter 4, the change in tuning can be one or two orders of magnitude slower. However, if there has been no change in algorithm and no change in tuning objective

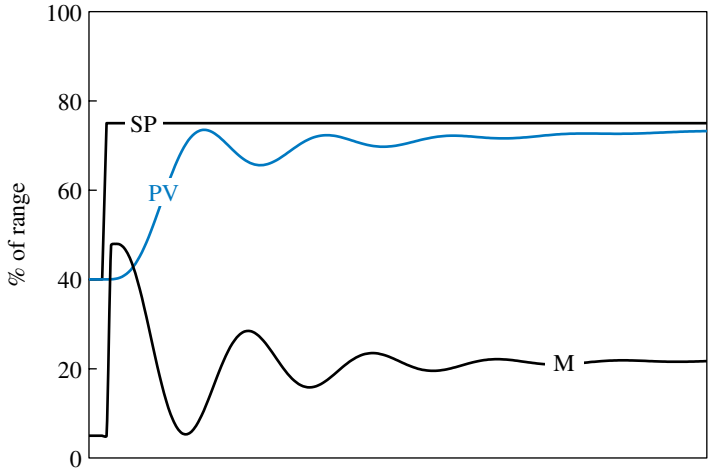

Figure 3.69 Kickback

then changes in K_c should be restricted to around 20% of the current value. Changes of around 50% may be made to T_i and T_d. The controller is tested with a SP change following each incremental change to the tuning and, provided it exhibits no problems, the next increment made.

This testing of controller tuning presents an opportunity to re-identify the process model. Provided a computer-based model identification technique is applied then SP changes made to validate the tuning can also be analysed to determine the process dynamics. Since the data are likely to be collected routinely by the process information system, this re-evaluation takes little additional effort.

Following the tuning method presented here should obviate the need for tuning by trial-and-error. However, the method does assume that the model dynamics have been determined accurately and that they are close to first order. This chapter would not be complete therefore without offering some guidance in this area. Controller gain affects all three P, I and D actions and should therefore be adjusted first. Steps of 20% are reasonable until the optimum value is approached, when smaller changes can be made. Adjustments to integral action can be made initially in much larger steps – either halving or doubling the action. If slightly oscillatory then controller gain may be reduced. Derivative action can be similarly adjusted. All three constants can then be fine-tuned to give optimum performance. In doing so the controller may show *kickback*. This is illustrated in Figure 3.69; the PV turns before reaching SP. This indicates that controller gain is too high and integral action is insufficient.

3.26 Tuning Cascades

We have already addressed the tuning of simple cascades in Chapter 2. One can argue that a cascade is not optimally tuned if the secondary is tuned independently of the primary. Indeed, it is often possible to apparently improve controller performance by simultaneously optimising the tuning of both primary and secondary. However, the likelihood is that the secondary will no longer function well as a standalone controller – often being far too oscillatory.

A more sophisticated approach may be justified. A problem can arise with cascade controllers when responding to load changes. Consider the double cascade arrangement shown in Figure 3.70. Added to our

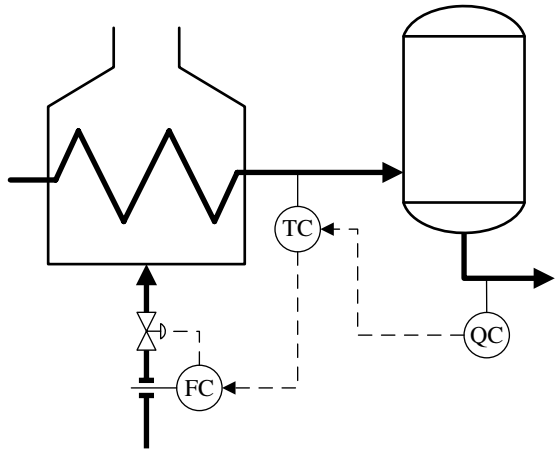

Figure 3.70 Double cascade control

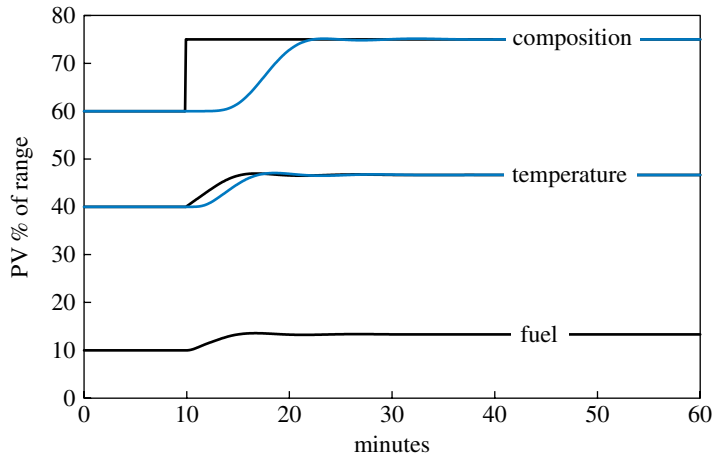

Figure 3.71 Response to SP change

fired heater is a reactor with its product composition dependent on reactor inlet temperature. All three controllers have been designed using the methods recommended in this chapter. Indeed, Figure 3.71 shows that they work well in response to an increase in the SP (black line) of the composition controller. However, as Figure 3.72 shows, when subjected to a load change, for example a reduction in fuel gas heating value, the composition controller is quite oscillatory. The temperature controller responds by increasing the fuel and will return the temperature to SP quite effectively. However, the transient disturbance in the temperature causes, sometime later, a disturbance to the composition which its controller attempts to correct. The oscillatory behaviour can be reduced by reducing the composition controller gain, but this is not ideal and will also unnecessarily slow down the response to SP changes.

In fact, the changes the composition controller makes to the temperature SP are largely unnecessary. The composition will ultimately return to that before the disturbance without any correction to the temperature SP. Figure 3.73 shows an enhancement to the controller to deal with this problem. It is based

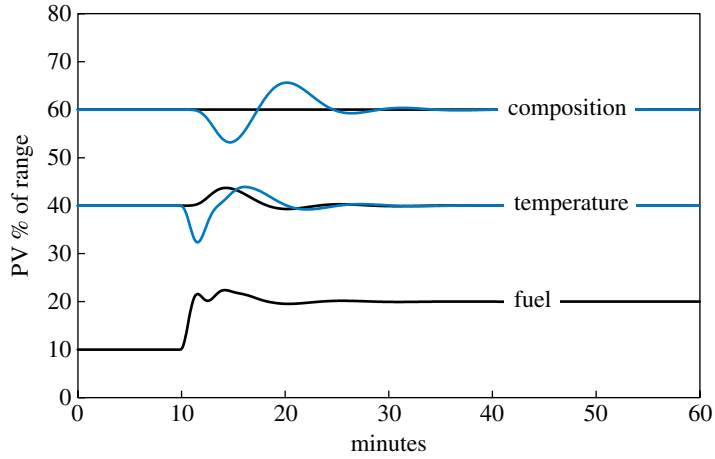

Figure 3.72 *Response to load change*

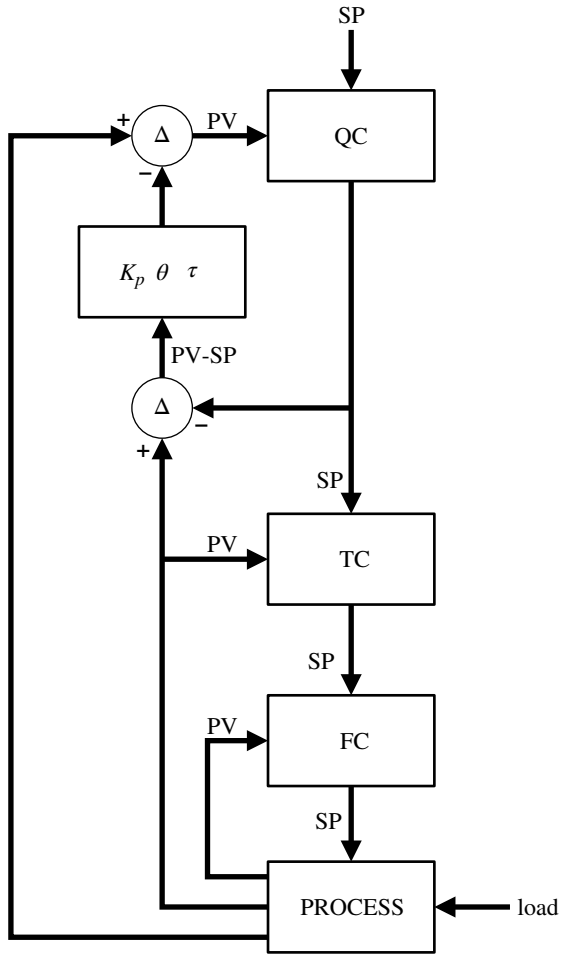

Figure 3.73 *Cascade compensation scheme*

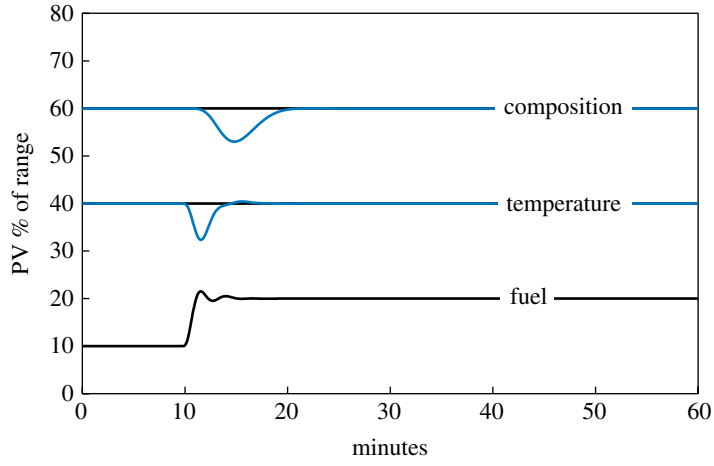

Figure 3.74 *Improved response to load change*

on a scheme proposed by Bartman [21]. The scheme is designed by determining the dynamic relationship between composition and temperature. We perform the test using the methods described in Chapter 2. However, unlike the dynamics required for controller tuning, we require the dynamics of the response to changes in temperature PV, not its SP. These dynamics (K_p, θ and τ) are applied to the temperature controller error. In effect, they predict how much of the deviation from composition SP is due to the deviation from temperature SP. This is subtracted from the measured composition used in the composition controller. If the dynamics are accurate and there is no other disturbance to the composition, then the composition controller will take no corrective action, as shown in Figure 3.74.

3.27 Loop Gain

In addition to process gain (K_p) and controller gain (K_c), a term often used is the *loop gain* (K_l). The loop gain is obtained by multiplying all the gain terms in the control loop. In the case of a simple PID controller, the loop gain is given by

$$K_l = K_p.K_c \tag{3.144}$$

We have seen in all the tuning methods that the product $K_p.K_c$ should be constant for a given value of θ and τ. Once we have established what the controller gain should be, we will need to change the value if there is any change which affects the loop gain. For example, since most PID controllers operate on a dimensionless basis, if the instrument range of either the PV or MV is changed, K_c will need adjustment. From Equations (2.2), (2.3) and (2.4) we define process gain as

$$K_p = \frac{\left(\dfrac{\Delta PV}{PV\ \ range}\right)}{\left(\dfrac{\Delta MV}{MV\ \ range}\right)} \tag{3.145}$$

So if we change the range of the PV or the MV, to keep the loop gain constant, the controller gain should be recalculated as

$$\left(K_c\right)_{new} = \left(K_c\right)_{old} \times \frac{PV\ \ range_{new}}{PV\ \ range_{old}} \times \frac{MV\ \ range_{old}}{MV\ \ range_{new}} \tag{3.146}$$

Note that by 'range' we strictly mean the *span* of the measurement. This is the difference between the *higher range value (HRV)* and the *lower range value (LRV)*. For example a temperature ranged from 100 to 300 would have a range of 200. Such a range, where real zero is below the minimum, is described as having a *suppressed zero*. If the measurement were ranged from –100 to 300 then its range is 400 and would be described as having an *elevated zero*. The *suppression ratio* is defined as the lower range value divided by the span. Our first example has a suppression ratio of 0.5; the second example has one of –0.25.

The same correction would be necessary if we change the MV of the controller, for example changing a primary controller cascaded to a flow controller so that it instead directly manipulates the control valve. This is often a 'quick fix' if the secondary flow transmitter has a problem. If the secondary flow controller has been well designed then, when the valve is fully open, the flow will be close to the flow transmitter range – in which case modification of controller gain will not be necessary. It is however important to check that this is the case – for example by using historical data to check that variation in flow (as percentage of range) is close to the variation in controller output. There may also be a significant change in linearity, previously dealt with by the flow controller, which must now be handled by the primary controller. Normally the dynamic response of the primary PV is likely to be similar for both changes in flow controller SP and in valve position – in which case there will be no need to change integral or derivative time. However, there will be situations where this is not the case. Consider control of a distillation column product composition using an on-stream analyser, cascaded to a tray temperature controller which in turn cascades to a flow controller. Omitting the temperature, by cascading the composition controller directly to the flow, will require significant tuning changes.

We will show in Chapter 6 that adding a ratio-based feedforward can also require recalculation of K_c again because the effective range of the MV may be changed.

3.28 Adaptive Tuning

Adaptive tuning, as the name suggests, automatically changes controller tuning constants as necessary to accommodate changes in process dynamics. One example is *gain scheduling* which changes the gain of the controller as the process gain changes. This may exist as a standard feature within the DCS or may require some custom coding by the engineer. The engineer may define some relationship between controller gain and process conditions – known as *scheduling variables*. This may be a table of values to be used as circumstances change or it may be some continuous function.

For example, one method of dealing with the highly nonlinear problem of pH control is to split the titration curve shown in Figure 3.75 into several linear sections as shown by the coloured lines. As the pH measurement moves between sections the controller would be configured to use a different process gain.

We showed in Chapter 2 that process gain (K_p) for most processes is inversely proportional to feed rate. To keep the loop gain constant the controller gain (K_c) could be scheduled to vary in proportion to feed rate – although we will show in Chapter 6 that there is a more elegant solution to this problem.

While the DCS vendor may describe the function as gain scheduling the technique can usually also be applied to the integral and derivative tuning constants.

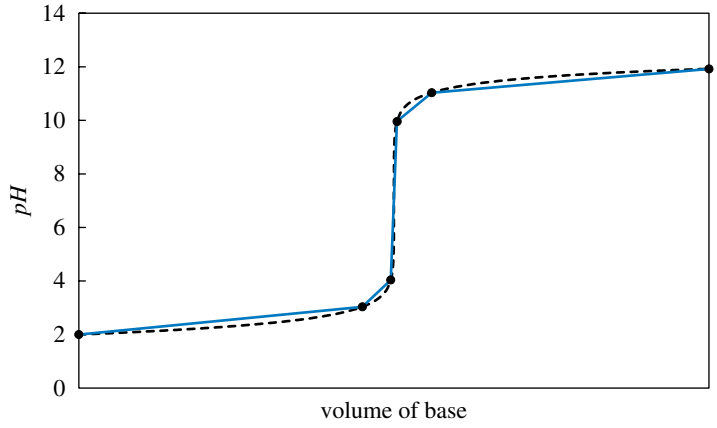

Figure 3.75 *Gain scheduling for pH*

In some cases it is possible to infer the process gain from process measurements. The controller is tuned for a known process gain $(K_p)_{design}$. This establishes the value for the loop gain. As the process gain varies the controller gain is automatically adjusted to maintain the loop gain constant.

$$K_c = \frac{\left(K_p\right)_{design}}{\left(K_p\right)_{measured}} \times \left(K_c\right)_{design} \tag{3.147}$$

It may similarly be possible to infer other process dynamics from process measurements. For example, deadtime is often related to residence time which in turn is related to the reciprocal of flow rate. Figures 3.42 and 3.43 show that both integral time and derivative time vary approximately linearly with deadtime. If deadtime could be inferred then T_i and T_d can be adjusted according to some simple formula. Such an approach is described as *programmed adaptive* control.

An approach described in some texts is *model reference adaptive control (MRAC)*. Here a reference trajectory is defined, in much the same way as λ is used in IMC tuning to specify the required response. If the PV deviates from this trajectory then controller tuning is adjusted to correct the deviation. It is, however, difficult to develop an update mechanism which results in stable robust control. In practice this approach is less likely to succeed than those described above.

There is a range of *self-tuners* on the market which attempt to perform on-line model identification and re-tune the controller as the process dynamics change. However, these can apply tuning methods that do not meet the key criteria that we have identified. Further they should not be seen as a replacement for developing a sound understanding of why and how the process dynamics are changing, or indeed if there is simply a problem with the instrumentation. With this knowledge it is likely that a more rigorous solution could be engineered.

3.29 Initialisation

Initialisation is the process that takes place when a controller is switched from manual to automatic mode. Its purpose is to ensure that the process is not disturbed by a sudden change in controller output. We first touched on this subject when converting the full position version of the control algorithm to its velocity

form. On initialisation the output of a full position controller must be matched to the current value of the MV. On older controllers this exercise was completed manually. However, with the velocity form, we have to ensure only that the incremental change made by the controller is zero. This is achieved by *PV tracking*. When in manual mode the SP is maintained equal to the PV so that, when the mode is changed to automatic, the previous and current errors are zero. Thus the controller output will be zero. Once in automatic mode the SP stops tracking the PV; the controller will respond to any process disturbance and the operator may change the SP as required.

Such tracking can occasionally cause problems when switching controllers to automatic. It is advisable when configuring a controller in a DCS to place upper and lower limits on SP to prevent the operator accidently entering a value that might otherwise cause an operating problem. With PV tracking it is possible for the SP to move outside this acceptable range and some DCS then prevent the controller being switched to automatic. The operator will first need to manually adjust the process until the PV, and hence the SP, move into range.

A technique similar to PV tracking is *output tracking*, used when initialising a cascade control scheme. If the primary is in manual mode and the secondary in automatic, the output of the primary tracks the SP of the secondary. When the primary is switched to cascade mode, tracking stops and the secondary's SP is then adjusted as required by the primary. Some systems offer an alternative method of slowly ramping the SP of the secondary to the output of the primary. Depending on the system, the engineer may be able to select the rate of ramping. The ramping function may be linear or may take the form

$$SP_n = P.SP_{n-1} + (1-P)OP_n \qquad \text{where} \quad 0 < P \leq 1 \tag{3.148}$$

This inserts a first order lag between the primary's output and the secondary's SP. (We will see in Chapter 5 that this is also the form of the noise reduction filter used by most DCS.) Setting P to zero removes the ramp function so that, on switching to automatic, the secondary's SP will immediately be set to the primary's output. Low values of P can therefore potentially cause a large disturbance to the primary PV resulting in unnecessary corrective action by the controller. Large values of P add a substantial lag between primary and secondary, requiring the primary controller to be less tightly tuned. Some systems disable the ramp function when the SP reaches the primary's OP. This removes the lag and results in the tuning then becoming unnecessarily slow. Ramping is therefore inferior to output tracking. It is included because tracking may not, under some circumstances, be possible. Consider the scheme shown in Figure 3.76,

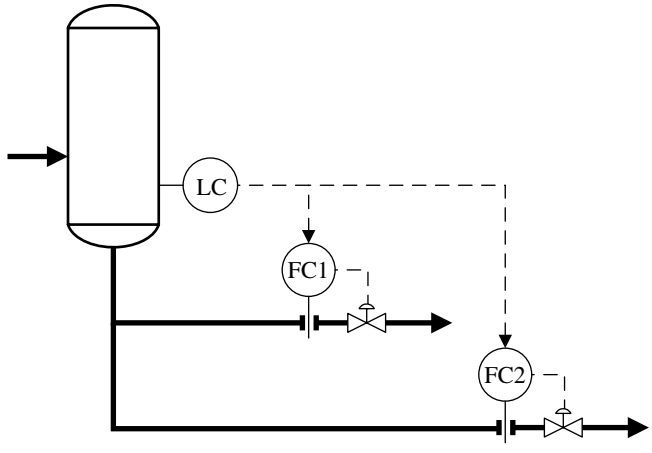

Figure 3.76 *Initialisation prevented by second secondary controller*

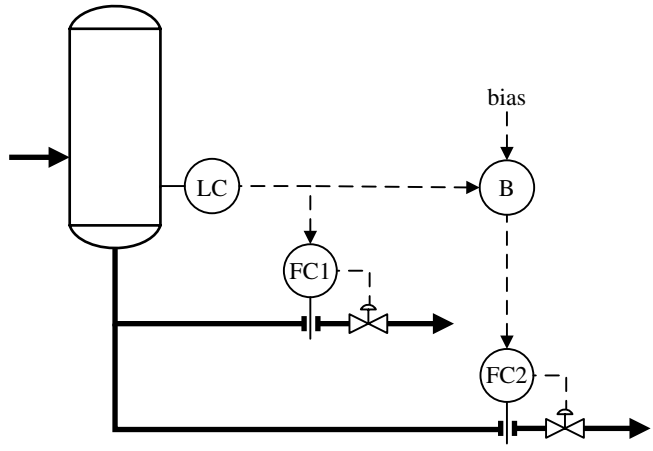

Figure 3.77 *Use of bias algorithm to permit initialisation*

where the output of the level controller is *fanned* to two flows. If the level controller is in manual mode then the SPs of the two flow controllers need not be equal. The level controller can only track one SP, for example that of FC1, and so SP ramping must be used for FC2.

It would be preferable, however, to modify the scheme to avoid the need for ramping. We will cover in more detail, in Chapter 6, the use of *bias* and *ratio* algorithms. A bias algorithm sums its two inputs, one of which will be the bias SP. A ratio algorithm multiplies its two inputs, one of which will be the ratio SP. Figure 3.77 shows the inclusion of a bias algorithm. It permits bumpless initialisation since, when in manual, the bias SP tracks the difference between the SPs of FC2 and FC1. We could alternatively have used a ratio; its SP would track the ratio of the SP of FC2 to that of FC1. With either, the additional initialisation path means that ramping would no longer be required. Since the operator can specify the bias (or ratio) SP, even when the level controller is in use, this approach also removes the requirement that the flows be equal.

3.30 Anti-Reset Windup

Control valves require *calibration* to convert the signal from the controller into a valve position. This calibration may be required in the valve positioner located in the field or in the DCS. The output range, coinciding to 0–100% valve position, may be 3–15 psi (in pneumatic systems) or (in electronic systems) 10–50 mA, 4–20 mA or 1–5 V. The actual output is usually permitted to move outside this range. Thus a valve not perfectly calibrated can still be driven fully shut or fully open. Similarly, if the valve is prone to stiction or hysteresis, it also overcomes the mismatch between position and signal that would otherwise prevent the valve from reaching its fully open or fully closed position. This is one reason why the ranges do not start at zero. Further it creates a distinction between a zero signal and loss of signal. Thus the controller output might vary from -25% to 125%, corresponding to a pneumatic signal of 0–18 psi or an electronic signal of 0–60 mA, 0–24 mA or 0–6 V.

There will be occasions when the controller *saturates*. For example, a flow controller may encounter a hydraulic limit so that, even with valve fully open, the SP cannot be reached. The integral (or reset) action will respond to this by continually increasing the output but, because the valve is fully open, will have no

effect on the flow. This is *reset windup*. Windup should be avoided because, if the process constraint is removed – e.g. by starting a booster pump, there will be a delay while the controller removes the windup and can begin actually closing the valve. This is resolved by keeping the permitted output range as narrow as possible, typically −5 to 105%.

However, the situation becomes more complex with cascaded controllers. The secondary can be controlling at SP but with its output at minimum or maximum. It is important therefore that the primary makes no changes to the secondary's SP which will cause it to saturate. DCS controllers have *external anti-reset windup protection*, sometimes described just as *external reset feedback*, to prevent this.

A similar technique is required with signal selectors. We cover these in more detail in Chapter 8, but a common use is to have two or more controllers outputting to a low or high signal selector. While one signal will pass through, the other(s) could potentially wind up. There must be logic in the selector that stops the deselected controller(s) from increasing their outputs (if routed to a low signal selector) or decreasing their outputs (if routed to a high signal selector).

3.31 On-Off Control

Before completing this section, it is right that we briefly examine the use of *on-off control* or (more generally) *two position control* – also called *bang-bang control*. While primarily used for temperature control in domestic systems (such as refrigerators, ovens, home heating, etc.) it does have some limited applications in the process industry. The technique is, in a heating application, to switch on the source of energy when the temperature is low and switch it off when the temperature is high.

Although the controller has a SP, there must be a *deadband* around this value within which no control action takes place. Without this the MV would be switched on and off at an unsustainable frequency. In domestic situations this deadband occurs almost accidentally as a result of the mechanism involved. Temperature is generally measured using a bimetallic strip which, as it bends, makes or breaks a contact. The distance it has to move between contacts provides the small deadband necessary. If this were not the case, then a deadband would need to be deliberately designed into the controller.

Figures 3.78 and 3.79 demonstrate the point. As the deadband is reduced, the PV will reach the edges of the deadband more quickly and so the frequency of MV switching increases. Reducing the deadband to zero would increase the frequency to the maximum the mechanics would allow – almost certainly soon

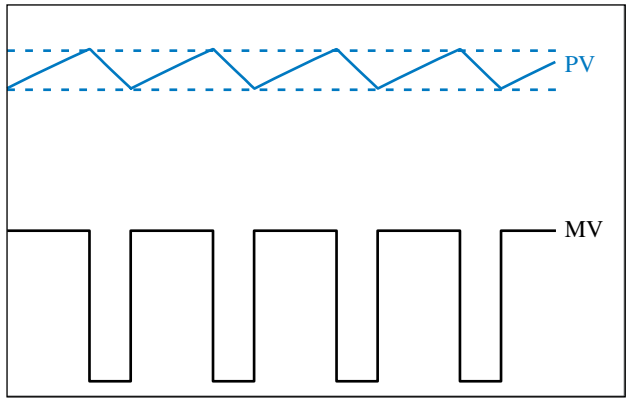

Figure 3.78 *On-off control with wide deadband*

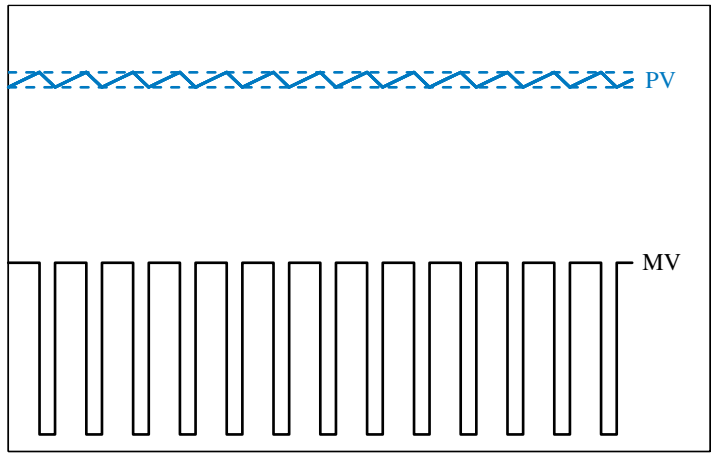

Figure 3.79 On-off control with narrow deadband

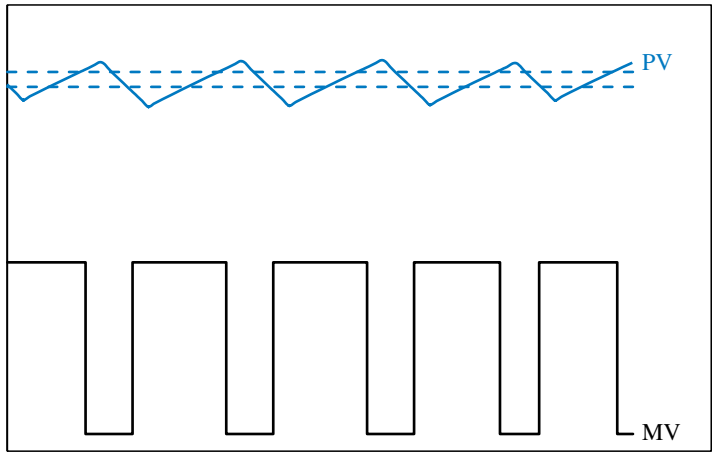

Figure 3.80 On-off control on process with deadtime

causing damage to the actuator. Note that here (and throughout this book) the term 'deadband' is quite different from that used by instrument manufacturers. They define it as the change in input that must take place, between reversals of direction, for there to be a perceptible change in output. It is a form of hysteresis introduced by the instrument.

In Figure 3.80 a small deadtime has been added to the process. Thus when the temperature reaches the high limit, even though the source of energy is switched off, the temperature continues to rise until the deadtime has elapsed. Similarly the temperature will fall below the low limit. Despite retaining the narrow deadband, the temperature deviates further from SP. On-off control is thus only applicable on industrial processes where tight control is not necessary and where deadtime is negligible. This restricts its use primarily to some level controllers. Typically it would be implemented using high and low level limit switches that would activate a solenoid valve. It can be emulated with a high

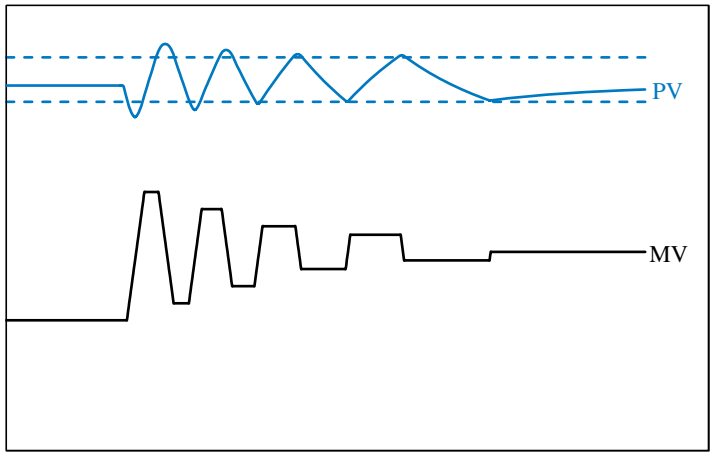

Figure 3.81 *Single speed ramp control*

gain proportional-only controller with a deadband, but care needs to be taken that excessive control action does not damage the control valve.

A modified form of on-off control, known as *single speed ramp control*, ramps the MV at a fixed rate when the PV is outside the deadband. If the ramp is much faster than the process dynamics then the MV limit will be reached before the PV moves inside the deadband. The response will then largely be that shown in Figure 3.78 but with the MV changing as ramps rather than steps. However, if the ramp rate is reduced sufficiently, the response will be as shown in Figure 3.81. In this example a load change has caused the PV to move outside the deadband causing the controller to take corrective action. Provided there are no more disturbances the process will eventually reach steady state. However, while used in some mechanical control systems, this form of controller offers no advantage in the process industry. Most control systems permit a deadband, or *gap*, to be configured within the standard PID algorithm. Properly tuned, the performance of this controller will be much the same. In Chapter 4 we cover a common application of gap control. A further refinement is to vary the ramp rate in proportion to the deviation of the PV from its SP. However, this results in the integral-only controller, as described by Equation (3.86), which is also a readily configurable option of the standard PID algorithm in most systems.

4

Level Control

So why do we dedicate a chapter to level control? What makes it so different from controlling other key process parameters such as flow, pressure and temperature? There are several reasons:

- The process behaviour is different. It is the most common example of a non-self-regulating (or integrating) process. It will not, after a change is made to the manipulated flow, reach a new equilibrium. The level will continue moving until either the process operator or a trip system intervenes. This affects the way that we execute plant tests and the way that we analyse the results.
- We may wish to apply very different tuning criteria. It may be more important to minimise disturbances to the manipulated flow than it is to maintain the level close to SP. This type of controller performance is known as *averaging* rather than *tight* level control. Averaging control can dramatically reduce the impact that flow disturbances have on a process.
- Most DCS offer a range of nonlinear algorithms intended to address specifically some of the problems that can arise with averaging level control. While of secondary importance compared to applying the correct tuning, they can be particularly useful in dealing with processes that experience a wide range of flow disturbances.
- Cascade control is usually of benefit but for reasons different from most other situations. Rather than offer the more usual dynamic advantage, it permits more flexibility in tuning and simplifies the calculation of tuning constants.
- While the use of filtering to reduce the effect of measurement noise affects the dynamic behaviour of any process, in the case of level control its impact is usually substantial and ideally should be avoided. Level measurements can be prone to noise – particularly in vessels where boiling is taking place.

4.1 Use of Cascade Control

Before tuning the level controller, we must decide whether it should act directly on the valve or be cascaded to a secondary flow controller, as shown in Figure 4.1. The general rule in applying cascade control is that the secondary should be able to detect and resolve any disturbance before the primary. Failure to adhere to this can result in instabilities caused by the secondary attempting to correct for a disturbance that has already been dealt with by the primary. Since the vessel level will change at almost the same time

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

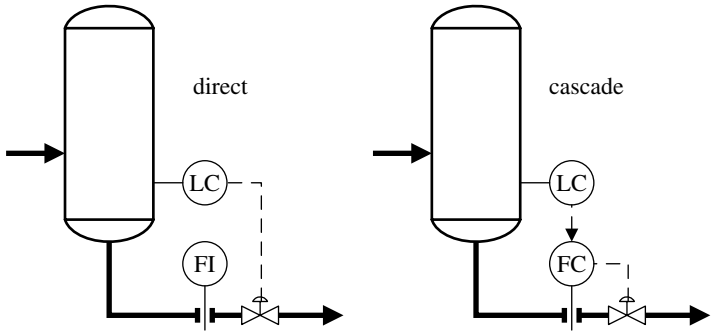

Figure 4.1 Alternative control configurations

as the flow, there would appear to be no dynamic advantage in applying cascade control. Indeed, this is the case if our objective is tight control. However, for averaging control, there is another consideration.

Imagine that a feed surge drum experiences an upstream or downstream fluctuation in pressure. The change in pressure drop across the manipulated flow valve will cause a change in flow. If this valve is under flow control then the disturbance will be dealt with quickly, resulting in little fluctuation to either the drum level or the manipulated flow. However, with no flow controller, the level controller is left to handle the disturbance. Since we want the manipulated flow to be as steady as possible, the level controller will need to be tightly tuned so that the control valve is moved quickly to compensate for the change in pressure drop. This is in conflict with the way we want the level controller to behave if there is a change in the uncontrolled flow. Under these circumstances we would want averaging level controller tuning. Applying a cascade allows us to meet both objectives. The flow controller would respond quickly to pressure changes, while the level controller would respond slowly to flow changes. A similar situation arises if liquid flows from the vessel under gravity, rather than being pumped. A flow controller will keep the flow constant despite any change in liquid head.

There is a secondary advantage to using a cascade arrangement when it comes to tuning both tight and averaging controllers. Both calculations require the range of the manipulated flow. This value is a constant if a flow controller is in place; without one, the range will vary with operating pressure and stream properties.

Orifice-type flow meters require a straight run length equal to 20 pipe diameters upstream and 10 downstream; they can be very costly to retrofit if this does not exist. The incremental cost of including the measurement in the original process design will be much smaller. If the construction budget is a constraint, the installation can be limited to the orifice flanges and orifice plate. The remainder of the instrumentation can be then be added if necessary later without incurring the cost of pipework modification.

The schematic of the process on which most of this chapter is based is included as Figure 4.2. This shows the level controller manipulating the discharge flow from the vessel. In this case, the inlet flow is the DV; the outlet flow is the MV and the level is the PV. However, there are situations where it is necessary to manipulate inlet flow. This makes no difference to the tuning calculations or controller performance – provided the engineer remembers to reverse the control action.

4.2 Parameters Required for Tuning Calculations

Level measurement is usually displayed as percentage of range, rather than in engineering units such as m or ft. In order to calculate controller tuning constants we first need to determine the working volume (V) of the vessel. This is the volume between 0% and 100% of the range of the level gauge. This can be

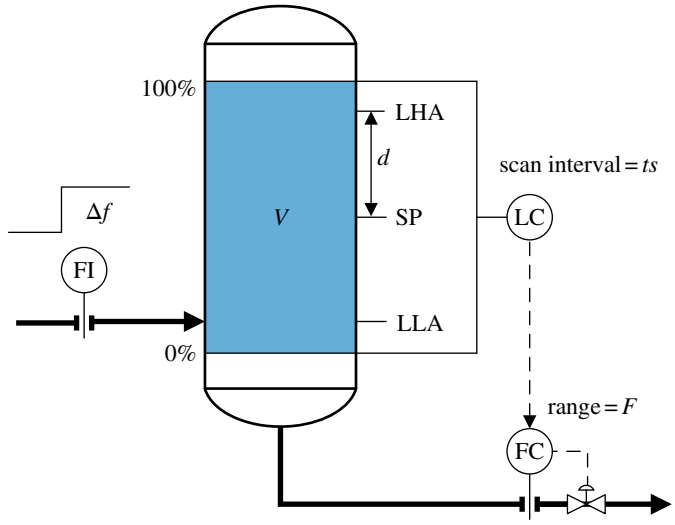

Figure 4.2 *Process flow diagram*

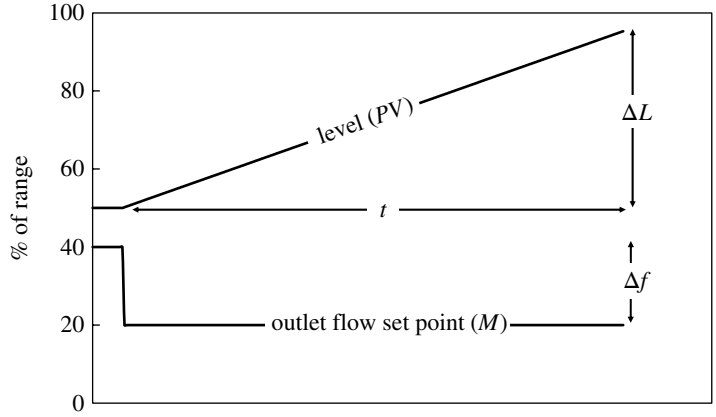

Figure 4.3 *Analysis of plant test*

determined by performing a simple plant test. Starting with the process at steady state, we decommission any existing level controller and step either the inlet or outlet flow to cause a flow imbalance (Δf). We allow this imbalance to exist for a known time (t) and record the change in level indication (ΔL). Of course, because the process is not self-regulating, we must end the test by restoring the flow balance before the level violates any alarms. The test result is shown in Figure 4.3.

We can then calculate the volume using Equation (4.1).

$$V = \frac{100\Delta f.t}{\Delta L} \tag{4.1}$$

Care should be taken with the engineering units. ΔL is in % (hence the 100 in the expression). The duration of the test (t) should be in units consistent with the flow imbalance (Δf). So, for example, if the flow

is measured in m^3/hr, t should be in hours. If in USGPM (US gallons per minute), then t should be in minutes and if in BPD (barrels per day), t should be in days.

This calculation assumes a linear relationship between volume and level indication. For vertical drums, assuming no large nozzles or internals, this will be the case. For horizontal drums and spheres the relationship is theoretically nonlinear but, providing the level gauge has been correctly ranged, the effect may generally be ignored. The larger the level change resulting from the test, the more representative will be the estimate of V.

Of course, if the vessel dimensions are known, it is possible to calculate the working volume. For a vertical drum, the calculation is trivial, i.e.

$$V = \pi r^2 \left(h_{100} - h_0\right) \tag{4.2}$$

The radius of the vessel is r, h_0 is the height of the 0% level indication (measured from the base of the vessel) and h_{100} is the height of the 100% level indication. Care should be taken in quantifying these values. They will usually not correspond to the location of the nozzles to which the level gauge is connected. The difference $(h_{100} - h_0)$ is the instrument range, usually found on the instrument datasheet. While a value for h_0 is not required for vertical drums, it is required for other shapes.

Again, care should be taken with units. If flow is measured in m^3/hr then r and h should be in m. If flow is in USGPM and r and h are measured in ft then a multiplier of 7.48 is required to convert ft^3 to USG. If the flow is in BPD then the multiplier should be 0.178.

In many cases the calculation of the working volume of a horizontal cylindrical can be approximated to the volume of a cuboid – multiplying its length by its diameter by the range of the level transmitter. More accurate calculation is complex. Firstly, we have to calculate the volume (V_0) between the bottom of the vessel and the 0% level indication. The length of the vessel (l) is that measured between *tangent lines* – where any dished ends are welded to the vessel. The last term in Equation (4.3) determines the volume of liquid held in the dished ends. It assumes a 2:1 ratio between drum radius and depth of each dish. It should be omitted if the vessel has flat ends.

$$V_0 = \left[r^2 \cos^{-1}\left(\frac{r - h_0}{r}\right) - \left(r - h_0\right)\sqrt{2rh_0 - h_0^2} \right] l + \frac{\pi h_0^2}{6}\left(3r - h_0\right) \tag{4.3}$$

The working volume (V) may then be derived (again omitting the term for the dished ends if not required).

$$V = \left[r^2 \cos^{-1}\left(\frac{r - h_{100}}{r}\right) - \left(r - h_{100}\right)\sqrt{2rh_{100} - h_{100}^2} \right] l + \frac{\pi h_{100}^2}{6}\left(3r - h_{100}\right) - V_0 \tag{4.4}$$

The same form of equation can be used to assess the linearity of the volume/height relationship. Equation (4.5) permits the measured volume (V_m) to be calculated as a function of h.

$$V_m = \left[r^2 \cos^{-1}\left(\frac{r - h}{r}\right) - \left(r - h\right)\sqrt{2rh - h^2} \right] l + \frac{\pi h^2}{6}\left(3r - h\right) - V_0 \tag{4.5}$$

Similar, somewhat simpler, calculations can be performed if the vessel is spherical.

$$V_m = \frac{\pi}{3}\left[h^2\left(3r - h\right) - h_0^2\left(3r - h_0\right) \right] \tag{4.6}$$

More usefully, it is better to plot the function in a dimensionless form, i.e. % of working volume against % level indication, or

$$100\frac{V_m}{V} \qquad \text{versus} \qquad 100\frac{h-h_0}{h_{100}-h_0}$$

Figure 4.4 illustrates the impact of taking the unusual step of mounting the level gauge to operate over the full height of the vessel. As expected, the horizontal drum and sphere show significant nonlinearity. How level changes with volume determines the process gain. Figure 4.5 shows how this varies over the range of the level transmitter installed on a horizontal cylinder with dished ends. It shows liquid level (per volume of liquid) plotted against level indication. The minimum value of this ratio, for all vessels of this shape, is 0.0379. This occurs at 50% fill. If the process gain between rate of change of level and manipulated flow is to remain within ±20% of a fixed value, then the maximum that this level-to-volume ratio can be is 0.0379 × 1.5 or 0.0568. This would be achieved by taking the more usual approach of mounting the level gauge so that it does not operate over the full height of the tank. For example, as the dashed line shows, locating h_0 and h_{100} at 15% and 85% of the vessel height would result in acceptable variation in process gain. This would easily be accommodated by a well-tuned linear controller and, because of the nonlinearity, would sacrifice only 16% of the theoretically available capacity. Performing a similar analysis for a spherical vessel would show that h_0 and h_{100} should be set at 20% and 80% of the vessel height – sacrificing 21% of the theoretically available capacity.

It is equally possible through poor design, to greatly increase nonlinearity by poor siting of the level gauge. For example, locating h_0 close to the bottom of the vessel and h_{100} at around 25% of the vessel height would cause significant tuning problems. Figure 4.6 illustrates this.

The problem of nonlinearity is therefore best avoided at the vessel engineering stage. If the vessel is either intended to provide surge capacity, or will provide useful capacity – even if this is not its main purpose, then there are two main design criteria. The first is to position h_0 and h_{100} as far apart as possible without encroaching into any serious nonlinearity. This is to make maximum use of the vessel capacity. This seems obvious but it is very common for gauges of a very narrow range to be mounted on very tall

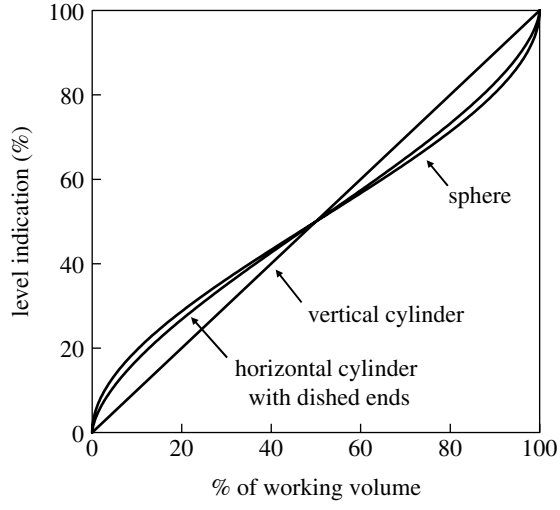

Figure 4.4 *Checking linearity of level indicator*

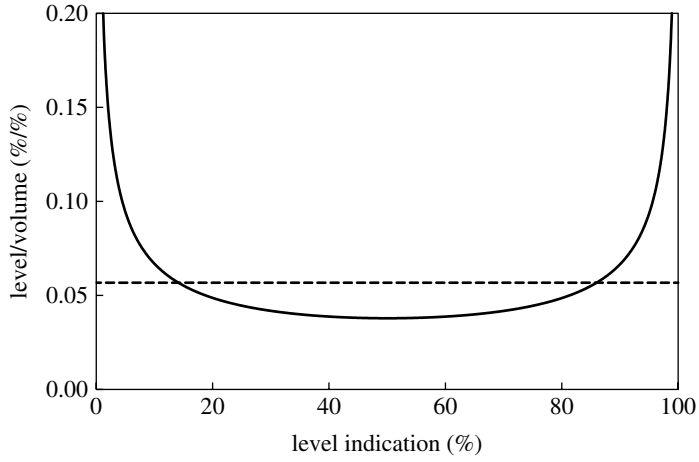

Figure 4.5 *Effect of liquid level on process gain*

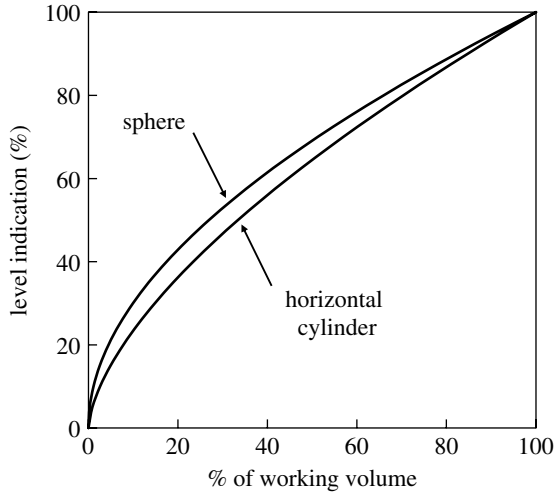

Figure 4.6 *Poorly placed level indicator*

vessels. The second, on horizontal cylinders and on spheres, is to position h_0 and h_{100} symmetrically either side of the centre line. The aim is to ensure there is equal capacity either side of the controller SP of 50%. The controller can thus handle equally both increases and decreases in flow. For example, if there is less capacity above the SP, the controller will need to be tuned for increases in the inlet flow and will not fully utilise vessel capacity when there is a decrease.

It is common for the control engineer to have to deal with problems inherent to poor design. Should nonlinearity present a problem, then this can be resolved with suitable signal conditioning. By definition, the level measurement (%) is given by:

$$L = 100\frac{h - h_0}{h_{100} - h_0} \tag{4.7}$$

Rearranging:

$$h = \frac{L}{100}\left(h_{100} - h_0\right) + h_0 \tag{4.8}$$

Substituting for h in Equation (4.5), or Equation (4.6) for spherical vessels, and building the resulting equation in the DCS will allow V_m to be continuously determined from L. Should the equation be too complex for the DCS, or use trigonometric functions that are not available, then an alternative approach is to fit a simpler curve of the form

$$V_m\left(\%\right) = a_1 L + a_2 L^2 + a_3 L^3 \tag{4.9}$$

The coefficients a_1, a_2 and a_3 would be derived from a series of volumes (as % of V) calculated by applying the exact equation off-line to a range of levels at (say) 1% intervals. For example, the curve for the horizontal cylinder with dished ends shown in Figure 4.4 can be described by setting a_1 to 0.3611, a_2 to 0.01907 and a_3 to –0.000126. This approach is sufficiently accurate between 5% and 95% of drum diameter. Should a gauge be installed over a wider range, the problem caused by any remaining nonlinearity will be insignificant to the process upset that has caused the drum to be virtually empty or full.

Some DCS support look-up tables which would permit the volumes for a number of selected levels to be entered directly. The system would then interpolate as necessary.

V_m may then be determined as a percentage of the working volume (V). Using this value as the measurement of the controller will present to the process operator a true measure of liquid inventory and changes in its value will be repeatable with respect to flow imbalances – no matter what the current inventory. This approach would be applicable to a vessel such as that shown in Figure 4.7.

The tuning methods we describe later result in very robust control design. Indeed, for averaging control, the value chosen for V has no effect on controller gain. Nonlinearity in the level/volume relationship therefore has to be quite severe before it causes noticeable degradation in control performance. The value of V obtained using Equation (4.1) will usually suffice. Indeed, it may well be more reliable than that calculated if vessel and instrument design records are out of date.

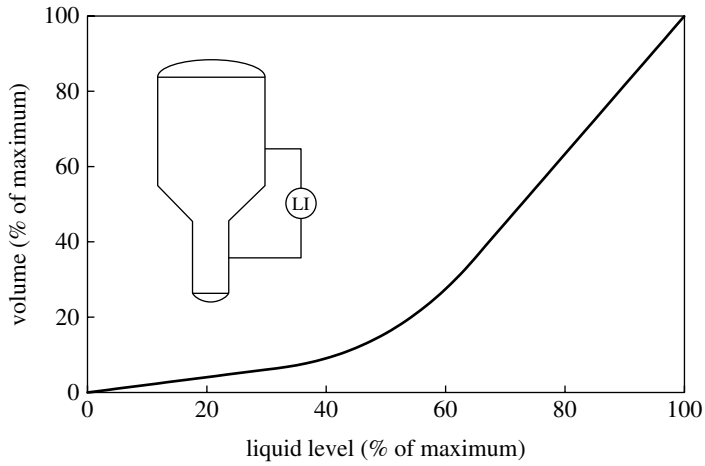

Figure 4.7 *Non-linear behaviour of level in a tapered vessel*

Other parameters required to permit controller tuning to be calculated are included in Figure 4.2. They include the normally expected flow disturbance (Δf). We will show later that this parameter has little effect on the design of tight level control, so choosing a precise value is unimportant. For averaging level control, some judgement should be used. If the plant is not yet commissioned, or historical data is not available for any other reason, then a value of 10% of the maximum flow is a good starting point. If process data do exist, then a visual inspection of flow trends (either inlet or outlet) should permit a sensible value to be selected. The important consideration is choosing a disturbance which is normal, i.e. the sort of disturbance we would expect not to generate any process alarms. We are not designing for a catastrophic reduction in feed due, for example, to equipment failure.

It is often the case that a process will largely experience minor flow disturbances but with the occasional larger upset. This might be caused by routine switches in process conditions such as drier swings, reactor regeneration, feed type change, change in operating mode, etc. If this is the situation (and we want to apply averaging level control), then two values for Δf should be chosen – Δf_1 for the small frequent disturbances and Δf_2 for the larger occasional upset.

Once the controller is commissioned, its performance should be closely monitored to confirm that the values chosen are realistic. For example, if surge capacity is not being fully utilised by an averaging level control, then a smaller value of Δf should be chosen and the controller tuning recalculated. The simplest method of doing this is to assess what fraction of the surge capacity is being used and then multiply the controller gain by this fraction.

The tuning method needs us to define how much of the vessel capacity may be used. This is set by the parameter d which is defined as the maximum deviation (in %) permitted from the level SP. Ideally, to make maximum use of surge capacity, this should be the distance between the SP and the nearest alarm. Placing high and low alarms symmetrically either side of SP will permit maximum use to be made of surge capacity. For tight control a much smaller value of d, e.g. 1%, would be selected.

Because controllers generally operate with their input and output in dimensionless form (e.g. % of range), we need the factor (F) to convert controller output into engineering units. If the level controller is cascaded to a flow controller then F is simply the range of the flow instrument. However, if the level controller acts directly on a valve, F is the flow with this valve fully open. If there is a flow measurement then F may be estimated by using historical data to correlate flow against valve position. Figure 4.8 shows a common situation. While the flow meter cannot be used in a secondary flow controller, it can be used to determine F. Figure 4.9 shows flow plotted against level controller output. Some of the scatter might be explained by the flow being measured on the inlet to the vessel while the valve is on the discharge. Much of this might be eliminated by excluding data where the level is not at SP. Another cause might be variation

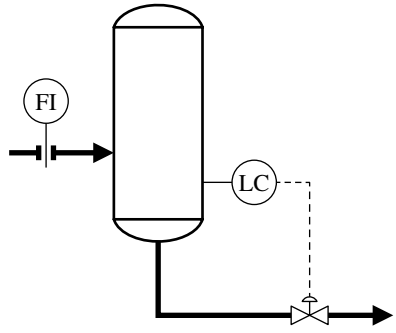

Figure 4.8 *Typical direct acting level controller*

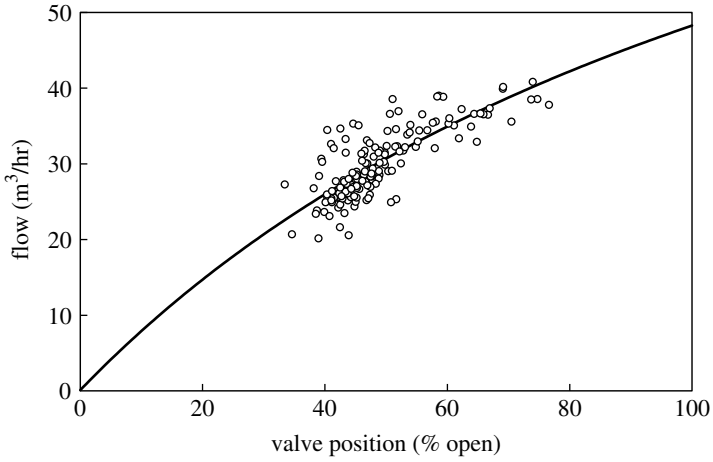

Figure 4.9 *Estimating range of flow control from historical data*

in upstream and downstream pressures. Valve hysteresis, as described in Chapter 2, might also be a problem. Nevertheless it is possible to fit a nonlinear valve equation, of the type covered in the next chapter, relating measured flow (f) to valve position (v %).

$$f = \frac{F.v}{100(1-k)+k.v} \tag{4.10}$$

In this example, the best fit is a value for k of 0.43 and that for F of 48.2 m^3/hr. If this approach is not possible then F may be approximated by multiplying the design flow by a factor of 1.3 – since this is typically the factor used in sizing the valve.

Finally, we need the level controller scan interval (ts).

4.3 Tight Level Control

Tight level control is required in situations where holding the level close to its SP is of greater importance than maintaining a steady manipulated flow. This would be applied, for example, to a steam drum level where we want to avoid the risk of routing water into the steam header and potentially damaging turbine blades. Similarly, on a compressor suction drum, we want to avoid routing to the compressor any of the liquid collected in the drum. As we shall see in Chapter 12, certain types of level controllers on distillation columns similarly require tight tuning. If reflux drum level is controlled by manipulating reflux flow then we must manipulate the overhead product flow to control product composition. This only has an impact because the drum level controller then takes corrective action and changes the reflux. In order for our composition control to act as fast as possible, the drum level controller must be tightly tuned. This would similarly apply to the level controller on the column base if it is set up to manipulate reboiler duty.

Controller tuning is derived by first assuming that we apply a proportional-only controller.

$$\Delta M = K_c\left(E_n - E_{n-1}\right) \tag{4.11}$$

Let us assume that before the flow disturbance, the level is at steady state and at SP, i.e. E_{n-1} will be zero. Since the flow imbalance (Δf) will have existed for one controller scan interval (ts), the current error (in dimensionless form) is given by

$$E_n = \frac{\Delta f.ts}{V} \tag{4.12}$$

In order to bring the level back to steady state, we need to restore the flow balance and so the controller must change the manipulated flow by the flow disturbance (Δf). In dimensionless form this means

$$\Delta M = \frac{\Delta f}{F} \tag{4.13}$$

The tightest possible control would be to take this corrective action in the shortest possible time, i.e. the scan interval (ts). By combining Equations (4.11) to (4.13) we can derive the largest possible controller gain (K_{max}).

$$K_{max} = \frac{V}{F.ts} \tag{4.14}$$

Care should again be taken with the choice of engineering units. Controller scan interval (ts) in most DCS is measured in seconds. So, if the flow range (F) is measured in m^3/hr, the result of this calculation should be multiplied by 3600 to ensure K_{max} is dimensionless. If the flow is in USGPM, then a factor of 60 is required. If the flow is in BPD, then a factor of 86400 should be used.

Examination of Equation (4.14) shows K_{max} is independent of Δf. This means that, no matter what size the flow disturbance, the controller will set the SP of the manipulated flow equal to the variable flow within one scan interval. Of course, control valve dynamics and the tuning of the secondary flow controller (if present) will mean the change in actual flow will lag a little, but nevertheless the controller should be effective.

Similar examination of the result shows K_{max} is dependent on ts. Unlike most controllers, a small change in scan interval (e.g. from 1 to 2 seconds) will have a dramatic effect on the required tuning.

Because the controller is proportional-only, it cannot return the level to its SP. However, the offset, given by Equation (4.12), will be extremely small and would probably not be noticeable – even if there are successive disturbances in the same direction as the first. If required, integral action may be added. To estimate how much, we first determine a vessel time constant (T) – measured with no controls in place. This is defined as the time taken for the level to change by the permitted deviation (d) following the flow disturbance (Δf). It is given by

$$T = \frac{Vd}{100\Delta f} \tag{4.15}$$

Since we require tight level control, we would select a very small value for d, e.g. 1%. Experience shows that, within a sensible range, the amount of integral action is not critical to controller performance. Empirically, setting T_i to $8T$ will give good control performance. Again care should be taken with engineering units. With Δf measured in m^3/hr, the result for T will be in hours. Although it is system-specific, the value of T_i is usually required in minutes and so a factor of 60 must be included. No factor would be needed if the flow is in USGPM. A factor of 1440 should be used if the flow is in BPD.

The additional control action introduced will mean that the controller will now over-correct. Compensation for the addition of integral action should be made by reducing proportional action. Again empirically, applying a factor of 0.8 to K_{max} works well. Derivative action is not normally beneficial to

level control – indeed, in the absence of any significant deadtime, even a small amount of action can cause instability.

Full controller tuning is therefore:

$$K_c = \frac{0.8V}{F.ts} \qquad T_i = \frac{V}{12.5\Delta f} \qquad T_d = 0 \tag{4.16}$$

The performance of a typical controller is shown as Figure 4.10. In this case the inlet flow was increased by 20% at the 8-minute point. The discharge flow was increased by the same amount in less than half a minute. Only the dynamics of the control valve prevented the correction being made more quickly. As a result the disturbance to the level would unlikely be noticed on a real process. Because the controller includes integral action, the discharge flow briefly exceeds the inlet flow in order to return the level to SP.

Some judgement is required before implementing the tuning suggested by Equation (4.16), particularly if far tighter than that already in place. The calculated tuning should be viewed as the tightest that should be applied – assuming no control valve lag and no measurement noise. The value of K_c derived is likely to be considerably greater than unity and will therefore amplify noise and may ultimately cause damage to the control valve. Controller gain may need to be reduced and slightly larger deviations from SP accepted. The use of filtering can be counter-productive. The filter will add significant lag to a process which likely only has control valve lag. The controller is likely then to be unstable and a large reduction in controller gain will be necessary to avoid this. This is illustrated in Figure 4.11. Noise is best dealt with at the vessel design stage. Turbulence in the vessel may be caused by the velocity of liquid entering the vessel through the inlet nozzle or, in the case of flash drums and steam drums, by boiling. The appropriate use of baffles and *stilling wells* will reduce the effect this turbulence has on the level measurement.

4.4 Averaging Level Control

Averaging level control is required in situations where keeping the manipulated flow as steady as possible is more important than keeping the level at its SP. Its aim therefore would be to make full use of the vessel capacity without violating any level alarms. Failure to appreciate the benefit of averaging level control, and how to design it, is one of the most common oversights in the process industry. There are many processes that would benefit from it greatly in terms of *disturbance rejection*.

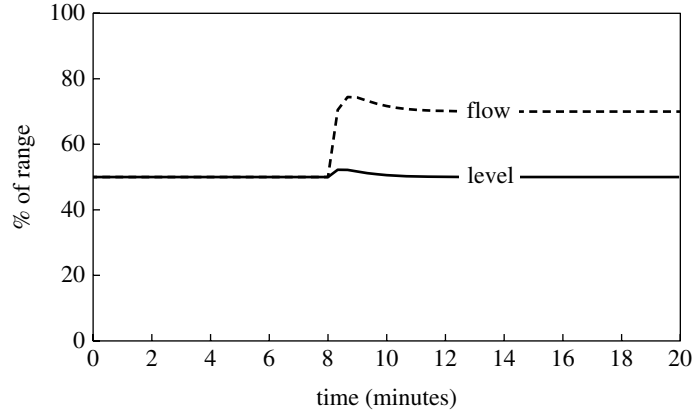

Figure 4.10 *Tight level control*

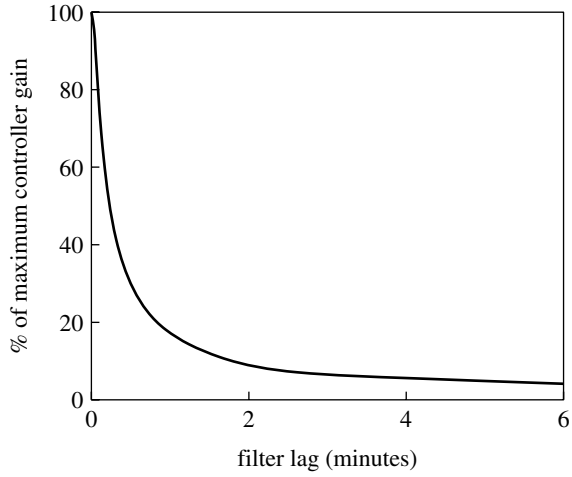

Figure 4.11 *Effect of filter*

The most obvious application is to feed surge drums. These are included in the process design specifically to reduce the effect of upstream flow disturbances on the downstream process. Installing tight level control in this situation makes the drum ineffective.

However, there are many situations where surge capacity is a spin-off benefit from a vessel that is in place for an entirely different purpose. For example, as we shall see in Chapter 12, it is common in a sequence of distillation columns for one column to be fed from the reflux drum of the preceding column. Provided the drum level controller manipulates the feed to the downstream column then averaging level control may be applied to minimise feed flow disturbances. Even if the overhead product is routed to storage, if the product is cooled by exchanging heat with another process stream, then disturbances to the energy balance can be reduced. The level controller at the base of the column may similarly be exploited if it is manipulating the flow of the bottom product. However, the available surge capacity may be small and therefore offer little opportunity. Also care must be taken if reboiler performance is affected by variations in level.

The main issue with averaging level control is its acceptance by the process operator. To achieve its objective, the vessel level will often approach alarm limits and may take several hours to return to SP. The operator may, not unreasonably, be quite concerned by this and not entirely persuaded that the benefit to the downstream unit is worth the apparent risk. A more cautious approach can allay such concerns. Initially tuning the controller to use only part of the available capacity and demonstrating over time that it does not violate this limit will help persuade the operator to accept use of all the available capacity – particularly if the benefit is demonstrable.

There are likely to be other related issues. Some sites permit the operators to configure process alarms; these will often then be set conservatively and the operator will need to be persuaded to relax them as far as possible. Asymmetry may be introduced; for example, there may be a problem with potential pump cavitation and therefore more concern about a reduction in level rather than an increase. The operator will raise level SP above 50% and possibly also increase the setting of the low-level alarm. This will mean full use is not made of the surge capacity when there is a flow increase. The converse may also apply, for example, if the operator is more concerned about overfilling the vessel.

The method used to tune the controller is very similar to that applied to tight level control. We start as before with a proportional-only controller. However, rather than eliminate the flow imbalance as quickly

as possible, we do so as slowly as possible. In this case the controller will take considerably more than one scan to make the correction, i.e.

$$\Delta M = K_c\left[\left(E_n - E_{n-1}\right) + \left(E_{n-1} - E_{n-2}\right) \ldots\ldots + \left(E_1 - E_0\right)\right] = K_c\left(E_n - E_0\right) \tag{4.17}$$

To make full use of the capacity we allow the level to reach the alarm just as steady state is reached. In other words we design for an offset of d, i.e.

$$E_n = \frac{d}{100} \tag{4.18}$$

By combining Equations (4.13), (4.17) and (4.18) we calculate the smallest possible controller gain (K_{min}).

$$K_{min} = \frac{100\Delta f}{Fd} \tag{4.19}$$

This, however, is just a first step in the controller design. Unlike tight level control we cannot retain such a proportional-only controller. As we can see in Figure 4.12, the level, as designed, remains at the alarm limit set at 90%. We will need integral action to return the level to its SP in preparation for the next disturbance. We determine this using the same method as for the tight controller. The full tuning then becomes

$$K_c = \frac{80\Delta f}{Fd} \qquad T_i = \frac{Vd}{12.5\Delta f} \qquad T_d = 0 \tag{4.20}$$

Wade [22] develops a tuning method by deriving, from first principles, a second order dynamic model of the process with its controller. While the method gives an unnecessarily oscillatory response for tight level control, for averaging control (substituting our terminology for that used) it gives a very similar result.

$$K_c = \frac{74\Delta f}{Fd} \qquad T_i = \frac{Vd}{18.5\Delta f} \qquad T_d = 0 \tag{4.21}$$

From this same dynamic model the process behaviour can be predicted. For instance, the time from the disturbance until the maximum deviation from level SP, known as the *arrest time*, will be $0.5T_i$. The MV overshoot will be 14% and this will occur at time T_i after the disturbance.

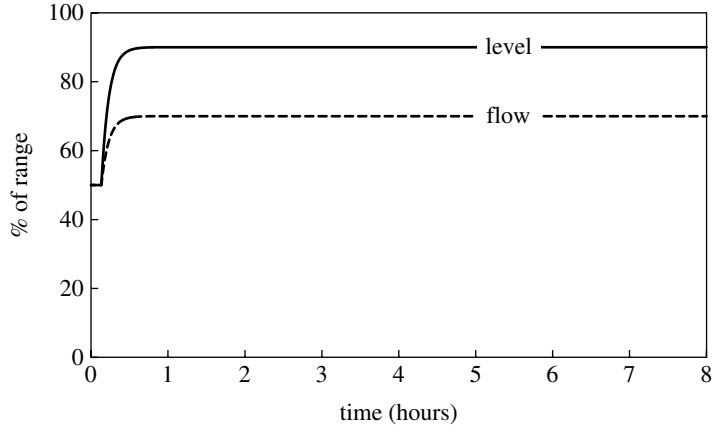

Figure 4.12 *Interim proportional only control*

Unlike the tuning calculations for tight control, calculation of K_c does not require the working volume of the vessel (V). Since the derivation of T_i is based on an empirical method a precise estimate of V brings little advantage. The impact of small inaccuracies, caused by nonlinearity or approximations made in calculating V is likely to be unnoticed – particularly as Δf will not be a constant.

Figure 4.13 shows how this controller would respond to the flow disturbance Δf with a SP of 50% and high level alarm at 90%, i.e. d is set at 40%. The uncontrolled flow was increased as a step change. As can be seen, the manipulated flow was increased as slowly as possible without violating the alarm. This increase took about 30 minutes, compared to the almost instantaneous increase that was made by the tight controller – substantially stabilising the downstream process.

While it is common for the concept of averaging control to be published, it is often the case that other, less easy to apply, tuning methods are presented. For example, it is possible to apply the Lambda tuning method as shown in Table 3.7. For most cases we can assume that, for level control, the process deadtime (θ) is zero. For each of the three algorithms, working with an integrating process, the Lambda tuning method gives

$$K_c = \frac{2}{K_p \lambda} \qquad T_i = 2\lambda \qquad T_d = 0 \tag{4.22}$$

In Chapter 2, we saw how K_p could be determined for plant testing. We will show later in this chapter (Equation 4.51) that it can also be derived from the vessel volume and flow range. Figure 4.14 shows the effect of varying λ. While generally understood to be the time constant of the trajectory of the PV following a change in SP, in the case of a level controller, it is also the time taken for the manipulated flow to change by the same amount as the flow disturbance. It is thus the time taken for the level to reach its maximum deviation from SP (the arrest time). While it is quite possible to choose a value of λ that gives effective averaging level control, as with self-regulating processes, selection is by trial-and-error. However, by combining Equations (4.20) and (4.22) for K_c with Equation (4.51), we can obtain a good starting value. As usual, care must be taken in working in consistent units of time.

$$\lambda = \frac{Vd}{40\Delta f} \tag{4.23}$$

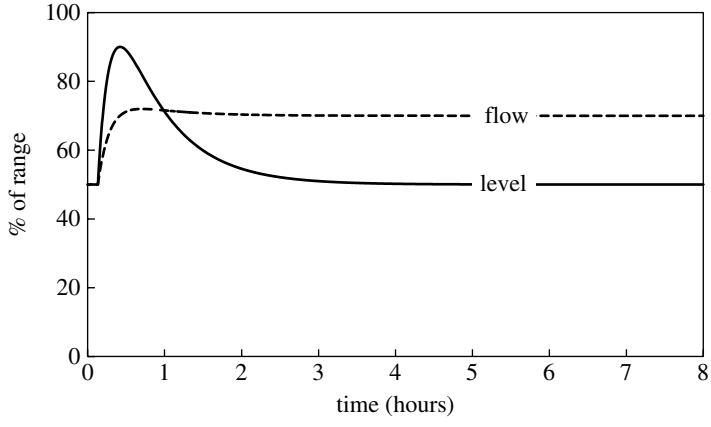

Figure 4.13 *Performance of averaging level control*

In the same way that we designed tight level control to eliminate the flow imbalance within one scan interval, we can specify a longer time in order to provide averaging control. Indeed, this is the basis of some of the proprietary level control algorithms. By defining t_h as the *time horizon*, we can modify Equation (4.16) to partly replicate the performance of these algorithms by applying the tuning

$$K_c = \frac{0.8V}{F.t_h} \qquad T_i = 8t_h \qquad T_d = 0 \tag{4.24}$$

If t_h is chosen to be T, as defined by Equation (4.15), then this tuning is the same as that derived from first principles.

Figure 4.15 illustrates how tuning constants, derived from Equation (4.20), vary on a typical surge drum as the maximum deviation (d) is changed from 1% to the maximum of 50%. Remembering that integral action is governed by the ratio K_c / T_i, the change in tuning moving from tight to averaging is more than three orders of magnitude.

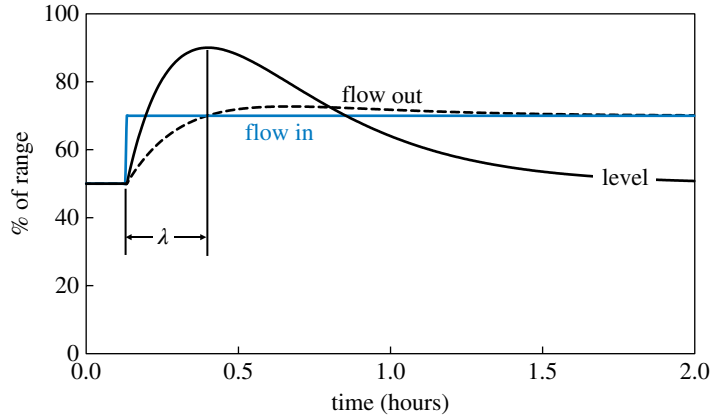

Figure 4.14 *Use of Lambda tuning method*

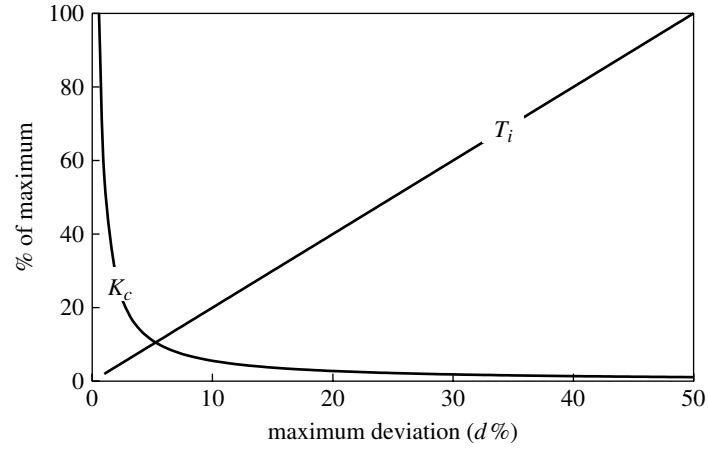

Figure 4.15 *Effect of permitted deviation on tuning constants*

Equation (4.20) can produce a value for T_i which is larger than the maximum supported by the DCS. Under these circumstances one of two approaches may be taken. The first is simply to set T_i to the maximum that the system will support and accept that full use will not be made of the available surge capacity. Clearly, whether this is effective will depend on how much greater the ideal value is compared to the maximum.

The alternative approach is to apply a proportional-only controller. Because flow disturbances will cause an offset, we need to ensure that the offset never violates an alarm. Rather than use the normal disturbance to determine the controller gain, we must instead use the minimum and maximum flow. The controller is designed so that the level will be at SP when the flow is midway between these values. The level will be at the low alarm at minimum flow and at the high alarm at maximum flow. The most conservative design basis is to assume the minimum flow is zero and the maximum is F. The maximum deviation from the mean flow is thus $F/2$. Replacing the normal disturbance Δf in Equation (4.19) with this value gives

$$K_c = \frac{50}{d} \tag{4.25}$$

A tuning technique often quoted is to set the proportional band (PB) of a proportional-only controller equal to the distance (in % of range) between the highest and lowest acceptable levels. Rearranging Equation (4.25) confirms that this technique is valid, provided that the SP is set midway between the high and low limits.

$$PB = \frac{100}{K_c} = 2d \tag{4.26}$$

For the controller to keep the level between limits, it is important that it is initialised so that controller output is 50% when the level is at SP. If the controller is switched to manual, returning it to automatic requires the operator to first set the output to 50% and, after switching the controller to automatic, change the SP to its normal value. One limitation of the proportional-only approach is that, if the SP is changed from its design value, the controller may not be able to resolve a violation of the maximum level deviation.

Whether a proportional-only controller is a more effective solution than using the maximum value of T_i will depend on the pattern of flow disturbances. If the minimum and maximum flows are only approached rarely then the full surge capacity will not be used. This is particularly true if Δf is small compared to the range of flow variation. Figure 4.16 compares the performance of the proportional controller to the PI controller, in terms of the change made to the manipulated flow. The proportional controller, since it must have a larger gain, initially changes the flow more rapidly. The PI controller must increase the flow above the steady-state value in order to bring the level back down to SP, but the overshoot is small and can be reduced further if necessary by increasing T_i.

Remember that if a proportional-only controller is configured as proportional-on-PV, it will not respond to changes in SP. This might be considered advantageous since it prevents the operator changing the SP to a value where the offset violates an alarm. However, it might create problems with operator acceptance, in which case the proportional-on-error algorithm can be used. Operators may also resist the use of a controller which deliberately results in an offset – arguing, for example, that a level kept low means that there is less inventory to deal with a temporary loss of feed. This argument can be countered because, if the level has reached the low limit, it has done so because the flow is also at its minimum. The next flow disturbance will therefore most likely be an increase. With the level very low there is the maximum capacity in the vessel to absorb this disturbance. Similarly if the level is at its high limit, this inventory can help alleviate the effect of the likely downward disturbance to flow.

One notable difference between the calculation for averaging control and those for tight tuning is the omission of *ts* from the calculations. Changing controller scan interval has no effect on controller tuning. However, unlike tight control, K_c is now strongly dependent on Δf. This begs the question as to how the controller will handle disturbances that are different from design. Figure 4.17 shows that a disturbance 25% larger than design causes an alarm violation – almost exceeding the instrument range. Similarly a disturbance 25% smaller results in underutilisation of surge capacity.

The simplest approach is to tune the controller based on the largest normally expected disturbance. This will avoid alarms but will underutilise surge capacity – a significant disadvantage if the larger disturbances are relatively rare. Under these circumstances a better approach would be to use a nonlinear control algorithm. Several different types of algorithm, specifically designed for averaging level control, are included in many DCS. While they can be tuned to give tight control, under these circumstances they offer no advantage over the normal linear version.

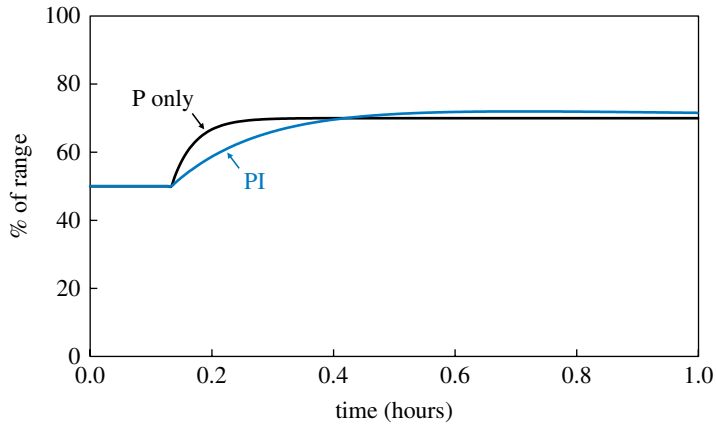

Figure 4.16 *Comparison between P only and PI control*

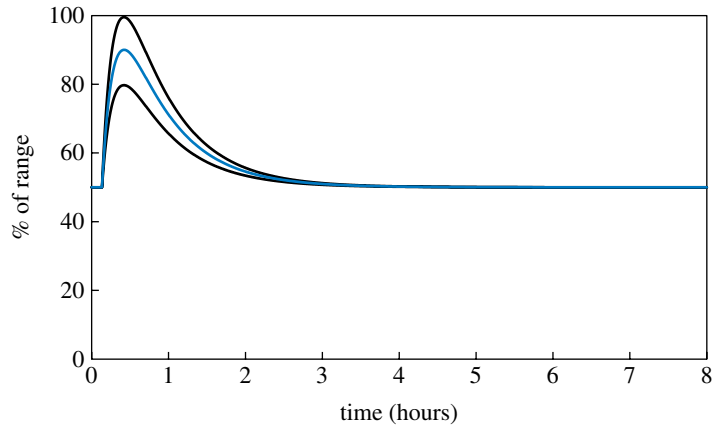

Figure 4.17 *Impact of size of flow disturbance on linear controller*

4.5 Error-Squared Controller

The most well-known nonlinear algorithm is *error-squared*. Strictly the error is not squared but multiplied by its absolute value, because we need to retain the sign if the error is negative. Since the controller works with a dimensionless error scaled between -1 and 1 (or -100% and 100%) the 'squared' error will have the same range. The effect is illustrated in Figure 4.18.

As is occasionally stated in some texts, error-squared is not intended to compensate for the nonlinearity between level indication and liquid volume in horizontal cylindrical drums (or spheres).

It is not usual to square each error term in the controller individually. The most common approach is to multiply the controller gain by the absolute value of the error. Omitting the derivative term (since we usually do not require this for averaging level control), the control equation becomes:

$$\Delta M = K_c |E_n| \left[\left(E_n - E_{n-1} \right) + \frac{ts}{T_i} E_n \right] \tag{4.27}$$

The effect of the additional $|E_n|$ term is to increase the effective controller gain as the error increases. This means the controller will respond more quickly to large disturbances and largely ignore small ones.

Tuning is calculated using the same approach as for the linear algorithm. We first determine K_{min} for a proportional-only controller based on restoring the flow balance when the offset has reached the alarm. In its continuous form we can write the control algorithm as

$$\Delta M = K_{min} \int_{t=0}^{\infty} E.dE = K_{min} \left[\frac{1}{2} E^2 \right]_{t=0}^{\infty} = K_{min} \left(\frac{1}{2} \left(\frac{d}{100} \right)^2 - 0 \right) \tag{4.28}$$

Combining with Equation (4.13) gives

$$K_{min} = \frac{100 \Delta f}{Fd} \left[\frac{200}{d} \right] \tag{4.29}$$

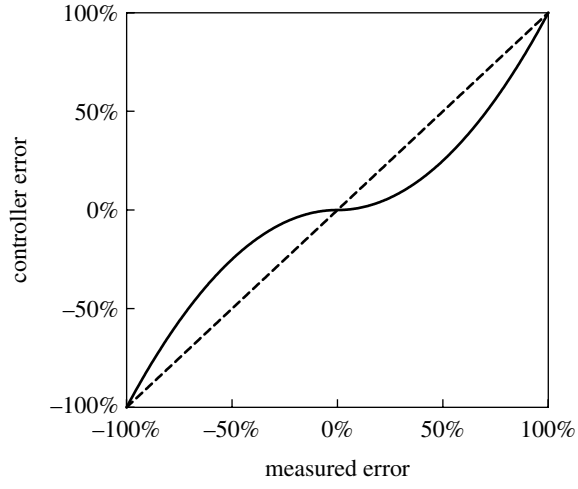

Figure 4.18 *Effect of error-squaring*

Following the same approach as the linear algorithm, the full tuning becomes

$$K_c = \frac{80\Delta f}{Fd}\left[\frac{200}{d}\right] \tag{4.30}$$

T_i and T_d are determined as in Equation (4.20). Figure 4.19 compares the performance of this controller compared to that of the linear version. It meets the design criterion of fully using the surge capacity without violating the alarm. However, it appears to show some oscillatory behaviour as the level returns to SP. The effect of error-squaring is to reduce the controller gain to zero when the error is zero. As the level returns to SP the small effective controller gain means that very little corrective action is taken and the level overshoots the SP. It is not until sufficient error accumulates that the controller gain increases enough for the flow imbalance to be reversed and the cycle then repeats itself.

In theory this oscillation will also be reflected in the flow. However, these changes will be almost imperceptible, having no effect on the downstream process. The changing level is a minor inconvenience. However, if noticed by an already reluctant process operator, it may cause difficulty in acceptance. And, if a real-time optimiser is installed, its steady-state detection logic may reduce the frequency of executions.

Fortunately a simple solution exists within most DCS. Rather than provide separate linear and error-squared algorithms the DCS will usually include a dual purpose algorithm. A typical example is:

$$\Delta M = K_c\left(C\left|E_n\right|+1-C\right)\left[\left(E_n - E_{n-1}\right)+\frac{ts}{T_i}E_n\right] \tag{4.31}$$

The additional term (C) gives the engineer the option of switching between algorithms. Setting C to 1 will give error-squared, while setting it to 0 gives linear performance. But the engineer is free to choose any value between these limits. By choosing a value close to 1, the controller will largely retain the nonlinear performance but the effective controller gain will no longer be zero as the error falls to zero. Controller tuning then becomes:

$$K_c = \frac{80\Delta f}{Fd}\left[\frac{200}{200\left(1-C\right)+Cd}\right] \tag{4.32}$$

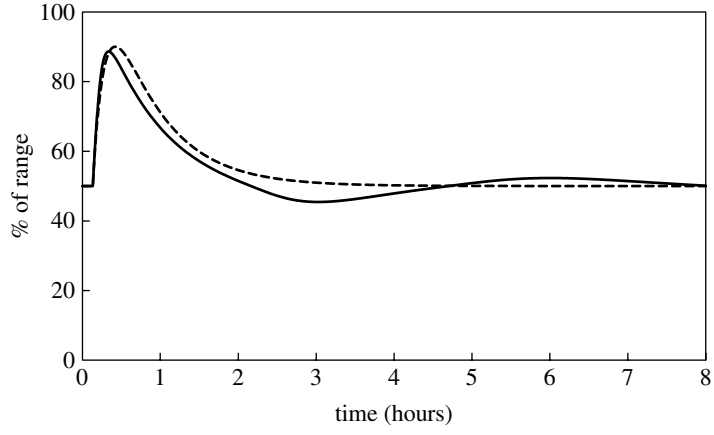

Figure 4.19 *Performance of error-squared versus linear*

T_i and T_d are determined as in Equation (4.20). Figure 4.20 shows the performance of this controller (with C set at 0.9) for the design disturbance and for disturbances 25% larger and smaller than design. The addition of the small amount of linear action has removed the oscillatory behaviour and, for the design case, given performance virtually identical to that of the linear controller. This algorithm, however, outperforms the linear controller for the non-design cases. Comparing the responses to those in Figure 4.17, for disturbances larger than design, the level violates the alarm by less and for a shorter period. For disturbances smaller than design, greater use is made of the surge capacity. While it does not completely solve the problem of varying flow disturbances it does offer a substantial improvement in performance.

It should be noted that the tuning calculation presented as Equations (4.30) and (4.32) are for the control algorithms exactly as described. DCS contain many variations of the error-squared algorithm. Even relatively minor changes to the algorithm can have significant effects on the required tuning. For example, squaring each error term individually appears to make a minor change to the integral action, i.e.

$$\Delta M = K_c \left[\left(|E_n| E_n - |E_{n-1}| E_{n-1} \right) + \frac{ts}{T_i} |E_n| E_n \right] \tag{4.33}$$

Comparing this to the controller described by Equation (4.27), the previous value of the error (E_{n-1}) is now multiplied by $|E_{n-1}|$ rather than $|E_n|$. Since the two values are measured only one scan interval apart, they will be almost identical and one would think this would have little impact on controller tuning.

Taking the same approach as Equation (4.17):

$$\Delta M = K_c \left[\left(E_n^2 - E_{n-1}^2 \right) + \left(E_{n-1}^2 - E_{n-2}^2 \right) ..... + \left(E_1^2 - E_0^2 \right) \right] = K_c \left(E_n^2 - E_0^2 \right) \tag{4.34}$$

Combining this with Equations (4.13) and (4.18), and applying the 0.8 factor, gives:

$$K_c = \frac{80 \Delta f}{Fd} \left[\frac{100}{d} \right] \tag{4.35}$$

Comparing this result to that in Equation (4.30) shows that a very minor change to the algorithm requires that the controller gain be halved to give the same performance. Other changes offered within some DCS include the option to apply error-squaring selectively to each of the proportional, integral and derivative

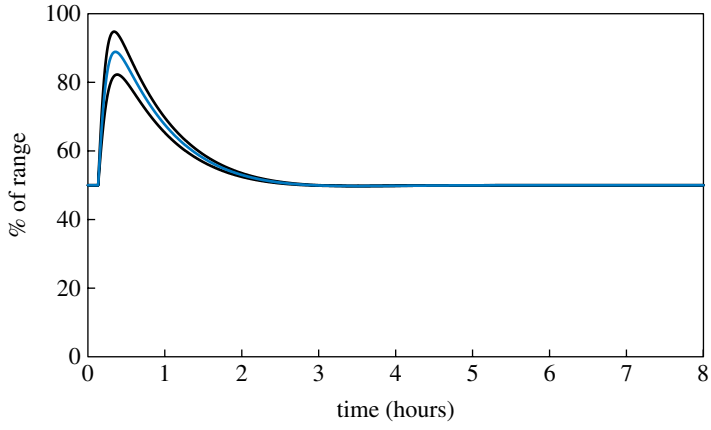

Figure 4.20 *Impact of size of flow disturbance on error-squared controller*

actions. There are also forms of the control algorithm that include other parameters to allow the engineer to specify the form of nonlinearity, for example:

$$\Delta M = K_c\left(C + K_n\left|E_n\right|\right)\left[\left(E_n - E_{n-1}\right) + \frac{ts}{T_i}E_n\right] \tag{4.36}$$

C may be set between 0 and 1. Setting it to 0 and the nonlinear gain term (K_n) to 1 gives the same form as error-squared algorithm as described by Equation (4.27). Similarly, setting C to 1 and K_n to 0 reduces the controller to the linear form. The controller described in Equation (4.31) can be emulated by setting K_n to $(1 - C)$. Some systems do not permit values for C other than 0 or 1. If the value we want to use is C^* (e.g. 0.9), then we must set C set to 1 and K_n to $(1 - C^*)/C^*$ (e.g. 0.111).

Controller tuning is determined from:

$$K_c = \frac{80\Delta f}{Fd}\left[\frac{200}{200C + K_n d}\right] \tag{4.37}$$

T_i and T_d are determined as in Equation (4.20).

4.6 Gap Controller

An alternative approach to introducing nonlinearity into the controller is to introduce a gap. In its simplest form this introduces a deadband around the SP within which no control action takes place. Outside the deadband the controller behaves as a conventional linear controller. The gap is configured by the engineer as a deviation from SP (G %). Tuning is given by:

$$K_c = \frac{80\Delta f}{F\left(d - G\right)} \tag{4.38}$$

T_i and T_d are determined as in Equation (4.20). Figure 4.21 shows the performance of this control with G set at 5% either side of SP. In this form it exhibits behaviour similar to that of the error-squared controller in

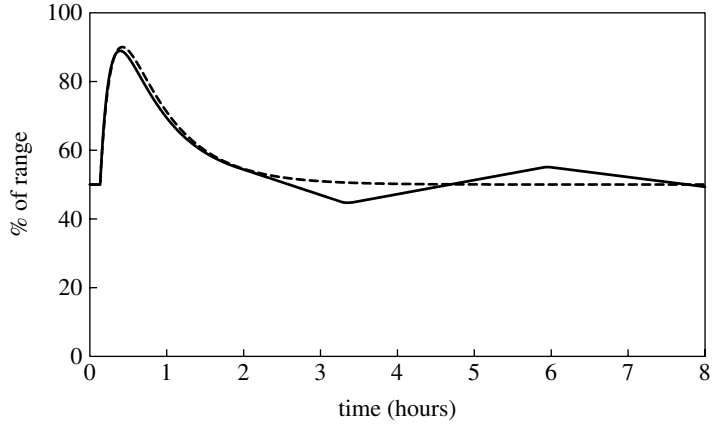

Figure 4.21 *Performance of gap controller with deadband*

that it will never settle at SP. Within the deadband no control action is taken and so any flow imbalance will be maintained until the level moves out of the deadband – at which point corrective action is taken to reverse the direction. While again this has little impact on the downstream process, it is undesirable for the same reasons as described for the error-squared controller, i.e. operator acceptance and steady state detection.

The solution is to apply a non zero gain within the gap. To preserve the required nonlinear behaviour, the value chosen should be substantially less than that used outside the deadband. Most DCS permit the engineer to define the value as a ratio (K_r). Where

$$K_r = \frac{\left(K_c\right)_{gap}}{K_c} \tag{4.39}$$

In which case the tuning is derived from:

$$K_c = \frac{80\Delta f}{F\left(d-\left(1-K_r\right)G\right)} \tag{4.40}$$

Typically K_r is chosen to be about 0.1, which will give performance much the same as the error-squared controller – including its ability to better handle non-design disturbances.

Alternatively, a value for $(K_c)_{gap}$ may be chosen and used in the following tuning method.

$$K_c = \frac{80\Delta f - FG\left(K_c\right)_{gap}}{F\left(d-G\right)} \tag{4.41}$$

The gap algorithm is better used in situations where flow disturbances can be classified into two types – relatively small changes (Δf_1) which take place frequently and much larger more intermittent changes (Δf_2). The controller is designed to deal with the smaller disturbance within the gap, thus

$$\left(K_c\right)_{gap} = \frac{80\Delta f_1}{FG} \tag{4.42}$$

The balance of the disturbance is then dealt with using the remaining vessel capacity, hence:

$$K_c = \frac{80\left(\Delta f_2 - \Delta f_1\right)}{F\left(d-G\right)} \tag{4.43}$$

Substituting Equations (4.42) and (4.43) into Equation (4.39):

$$K_r = \frac{\Delta f_1}{\Delta f_2 - \Delta f_1}\left[\frac{d-G}{G}\right] \tag{4.44}$$

Key to the performance of this controller is the choice of G. The same value should be used for both positive and negative variations from SP. This symmetry, combined with symmetrically placed high and low level limits, ensures that we do not have to tune the controller for disturbances in the more demanding direction and thus underutilise surge capacity for disturbances in the opposite direction.

For the gap to be beneficial, K_r must be less than 1. Applying this constraint to Equation (4.44) results in

$$G \geq d\frac{\Delta f_1}{\Delta f_2} \tag{4.45}$$

Applying a more realistic limit on K_r (e.g. 0.1) results in

$$G \geq d\frac{10\Delta f_1}{9\Delta f_1 + \Delta f_2} \tag{4.46}$$

The wider we make G, the smaller we make the drum capacity that the controller can use to deal with the larger disturbance. A larger controller gain will therefore be required. This gain given by Equation (4.43) should not exceed that required for tight control as determined by Equation (4.16). Thus

$$G \leq d - \frac{100\left(\Delta f_2 - \Delta f_1\right)ts}{V} \tag{4.47}$$

Between these constraints the choice of G is a compromise. Larger values will make better use of surge capacity during small disturbances but will leave little capacity to smooth larger flow changes. Again care should be taken with the choice of engineering units. T_i and T_d are determined as in Equation (4.20).

Figure 4.22 shows the performance of a well-tuned gap controller. In this case, G is set at 30% and the flow changed by Δf_1. The coloured line shows that the surge capacity is used as specified. The black line shows the result of a flow change of Δf_2, which is four times larger than Δf_1. The level deviation peaks at 40%, the value in this case for d. From Equation (4.44), we can see that K_r is set at 0.11.

If the DCS does not support gap control, then gain scheduling, as described in Chapter 3, could be configured to use different values of controller gain for different deviations of the level from SP. The gains would be determined from Equations (4.42) and (4.43). Such an approach would also allow more than two bands (and more than two sets of tuning) to be applied if required.

4.7 Impact of Noise on Averaging Control

The effect of measurement noise on averaging level control is somewhat different to its effect on tight control. Transmission of noise to the control valve is less likely to be a problem because the controller gain is substantially smaller. Nor is it likely that introducing a filter and its associated lag will give stability

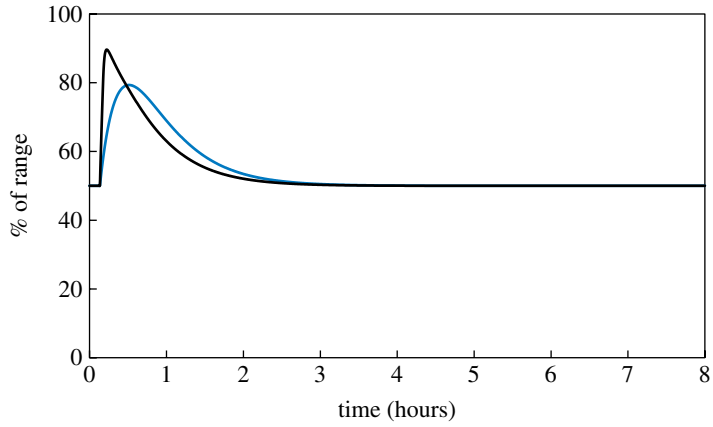

Figure 4.22 *Performance of gap controller with very different disturbances*

problems. However, when using the full surge capacity, as the level approaches alarm limits, the noise will cause nuisance alarms. This can be avoided by reducing the value chosen for d to take account of the noise amplitude. This will increase K_c and reduce the use of surge capacity. Filtering will not eliminate the need to increase K_c. It will have no effect on alarms unless the filtered value is used by both the controller and the alarm. The lag introduced will delay the response of the controller. As the size of the flow disturbance approaches the design value, the true level will violate the alarm before the controller can complete its correction.

It might be thought that nonlinear controllers deal better with noise and might therefore be considered for tight level control, where the high controller gain would otherwise amplify the noise. In theory, for small disturbances, the effective controller gain is small and hence little noise will be passed to the manipulated flow. However, such controllers, to compensate for the little action taken at the beginning of a disturbance, require a gain higher than that for a linear controller. This means that, as the level moves away from SP, noise amplification will become worse than that from a linear controller.

With controllers that are nonlinear over the full range of error, such as error-squared, noise can cause oscillatory behaviour. Different gains will be applied to negative and positive spikes of noise – so the average output from the controller will be different from that if there was no noise. This is illustrated in Table 4.1. It shows the situation where, at time = t, the SP is changed from 60% to 50%. A spike of noise, of ±1% around the PV, then occurs over the next four controller scan intervals. Since the algorithm is proportional-on-PV, the proportional action in response to the change in SP should be zero. But the noise causes proportional action of $0.02K_c$. And the integral action causes a change $0.02K_c.ts/T_i$ larger than it would be without noise. While small, these changes will be repeated for every noise spike. They speed up the return to SP and can trigger cyclic behaviour.

This is illustrated in Figure 4.23, where noise of ±1% of measurement range has been added to the example of the error-squared controller shown as Figure 4.19. The controller still responds well to flow changes but as it returns to SP the nonlinearity appears to amplify the noise to something in excess of ±10% – despite the controller gain approaching zero. In fact, the combination of noise and nonlinearity is triggering an oscillation with a period of about two hours. The frequency of oscillation is too low for there to be any noticeable impact on the manipulated flow, so the controller is still meeting the objective of maintaining this as steady as possible. The issue is that, because the oscillation is not a 'clean' sine wave, it can easily be mistaken for the controller successfully making use of available surge capacity, where in fact it is reducing the available capacity.

Filtering the measurement does not solve the problem. The level of filtering required introduces a lag which increases the amplitude of the oscillation. The effect can be reduced by the use of the dual purpose

Table 4.1 *Effect of error squared control on noise*

Time	SP	PV	$\lvert E_n \rvert (PV_n - PV_{n-1})$	$\lvert E_n \rvert E_n$
t	50	60	0.00	1.00
$t + ts$	50	59	−0.09	0.81
$t + 2.ts$	50	60	+0.10	1.00
$t + 3.ts$	50	61	+0.11	1.21
$t + 4.ts$	50	60	−0.10	1.00
total			+0.02	5.02

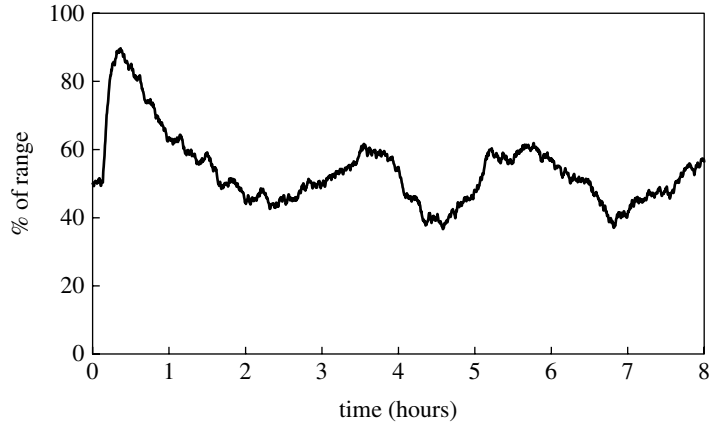

Figure 4.23 Performance of error-squared controller with noise

algorithm described in Equation (4.31) but, to eliminate it, C would have to be set close to 1 – almost removing the nonlinearity completely.

The better solution is to use a gap controller set up as described in the previous section. Since the nonlinearity only exists when the level crosses in or out of the gap then, for most of the time, the same gain is applied to both positive and negative spikes of noise. Indeed, if the gap controller is not required to handle very different flow disturbances, the deadband can be set at slightly larger than the noise amplitude so that noise is completely ignored when the level is close to SP.

4.8 Potential Disadvantage of Averaging Level Control

While averaging control is always beneficial in minimising disturbances to the downstream flow, consideration may need to be given to the impact it can have on process dynamics. This can be an issue if control of a PV downstream of the vessel is performed by adjusting an MV upstream. Commonly this situation can arise when the addition of a constraint controller, such as MPC, is proposed. These techniques are described fully later in Chapter 8, but consider now the constraint controller shown in Figure 4.24. It has

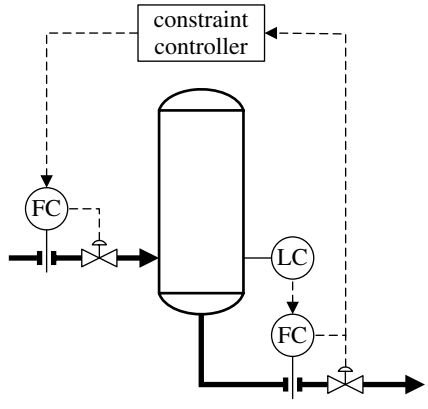

Figure 4.24 Impact of averaging control on process dynamics

been proposed because a downstream hydraulic constraint can result in the control valve being fully opened. To avoid this, the constraint controller monitors the downstream FC output and adjusts the SP of the upstream FC to maintain the process just inside the capacity constraint. The use of averaging level control would cause a large lag between the downstream valve position and the upstream flow SP – such that the constraint controller can only slowly correct for any deviation from target. Changing to tight level control would resolve this problem. The choice of LC tuning then becomes a compromise between two conflicting objectives.

The use of nonlinear level control algorithms can also present a problem. The dynamic relationship between the downstream valve position and the upstream flow SP will then vary – depending on the size of the deviation from the LC SP. This presents a problem in optimally tuning the constraint controller.

This example is of course rather simplistic. The addition of the constraint controller could be avoided by reconfiguring the level controller so that it manipulates the inlet flow – perhaps leaving the downstream valve fully open to truly maximise capacity utilisation. But this approach will not usually be practical for more complex MPC applications. The solution might be to remove the basic LC completely – moving its PV and MV into the MPC application.

4.9 General Approach to Tuning

So far we have adopted an approach to controller tuning which is specific to level control. It cannot be applied to other integrating processes such as some applications of pressure and temperature control. We have done this because conventional tuning methods do not readily lend themselves to averaging control or to nonlinear control algorithms. This does not mean that we cannot apply conventional methods to tight level control that use the linear algorithm. Indeed, we can predict the process gain that we would otherwise need to obtain from plant testing. Consider the general equation for an integrating process

$$PV = K_p \int MV.dt \qquad \text{or} \qquad \frac{dPV}{dt} = K_p.MV \tag{4.48}$$

We can write this (in dimensionless form) for our vessel:

$$\frac{dL}{dt} = K_p \frac{\Delta f}{F} \tag{4.49}$$

But we can predict the rate of change of level from the working volume of the vessel:

$$\frac{dL}{dt} = \frac{\Delta f}{V} \tag{4.50}$$

Combining Equations (4.49) and (4.50) enables us to predict the process gain. If F is measured in m^3/hr and V in m^3, then

$$K_p = \frac{F}{V}\text{hr}^{-1} = \frac{F}{60V}\text{min}^{-1} = \frac{F}{3600V}\text{sec}^{-1} \tag{4.51}$$

Unlike self-regulating processes, where K_p can be expressed in a dimensionless form, it now must have units of reciprocal time. Conventionally reciprocal minutes are used. Further Equation (4.51) shows that, again unlike self-regulating processes (discussed in Chapter 2), the process gain on an integrating process does not vary with feed rate.

The reciprocal of K_p is sometimes described as the *residence time* of the vessel. However, this may not be the same value as that quoted by the process design engineer since it only takes into account the liquid volume between 0 and 100% of the level gauge and assumes the flow is 100% of its instrument range.

To calculate tuning constants, using the methods described in Chapter 3, we also need the process deadtime (θ). For most level controllers this will be small. Choosing a value of a few seconds will result in controller tuning that will give a performance similar to tight control designed by applying Equation (4.16).

Combining the values K_c and T_i, for averaging control, from Equations (4.20) and (4.51) we obtain:

$$K_c T_i = \frac{80 \times V}{12.5 \times F} = \frac{6.4}{K_p} \tag{4.52}$$

Based on this equation some texts suggest if K_c is increased, for example to reduce the maximum deviation from SP, then it is important for T_i to be reduced in proportion. While in principle this is true, the value 12.5 was established empirically. Stable control will be achieved over a wide range of values for $K_c T_i$. For example, from Equation (3.112), Ziegler and Nichols [1] suggest $K_c T_i$ should be $3/K_p$. The Lambda tuning method (Table 3.7), for a PI controller on an integrating process with no deadtime, suggests $4/K_p$. Optimally tuning a PI controller for load changes (Table 3.8) suggests, if the deadtime is zero, a value of $6.7/K_p$ – very close to that established empirically.

This robustness is particularly helpful if gap control is employed. We would otherwise have to automatically adjust T_i as the gap is entered or exited. Similarly the use of the standard error-squared algorithm would otherwise require T_i to be divided by $|E|$. This effectively applies error-squaring to only the proportional action, explaining perhaps why this version of the algorithm is available in some DCS.

However, it is possible to apply this method to convert a tight level controller, with tuning derived from the process dynamics, to an averaging level controller. Figure 4.25 shows the effect of starting with the tuning used in Figure 4.10 and then repeatedly decreasing K_c. The response keeps the same shape but with the maximum deviation from SP increasing at each step. So, for example, to change from tight tuning that achieves a maximum level deviation of 1%, we could modify this to averaging tuning by dividing K_c by 40 and multiplying T_i by 40. This would give a maximum deviation of 40%. While not quite as effective as using Equation (4.20), as shown by the coloured trend, the difference in the stability of the downstream flow would likely be undiscernible.

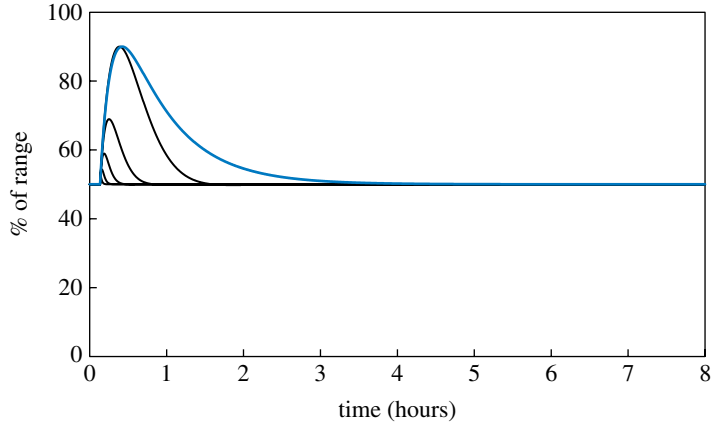

Figure 4.25 *Converting from tight to averaging level control*

There are level controllers that have substantial deadtimes. Consider the process in Figure 4.26. Level in the base of the distillation column is controlled by manipulating the reboiler duty. Unlike most level controllers it would be difficult (and probably unreliable) to predict the relationship between *PV* and *MV*. Further the reboiler introduces a large lag. The only practical way of identifying the process dynamics would be a plant test, as described in Chapter 2. The controller would then be tuned by applying one of the methods described in Chapter 3. This, unlike most level controllers, is likely to benefit from the use of derivative action. However, whether this level control strategy should be selected requires careful consideration. The process dynamics will restrict how tightly the level can be controlled without becoming unstable. While in some cases there may be no practical alternatives, its slow response to disturbances may restrict the performance of other controllers. For example, product composition would be controlled by adjusting the SP of the bottoms flow controller shown. It is unlikely that the LC could cope with rapid changes to this flow, or to the reflux flow, and correction of off-grade composition could only take place slowly. Full details of alternative approaches are given in Chapter 12.

4.10 Three-Element Level Control

Three-element level control is most commonly applied to the control of water level in steam drums on boilers. However, it is applicable to many other situations where tight level control is required and is made difficult by unusual dynamics.

The first most commonly encountered problem is *swell*. The water in the steam drum contains vapour bubbles which expand if the pressure in the drum is reduced, thus increasing the water level. So, if there is an increase in steam demand which causes a transient drop in drum pressure, the level controller will reduce the flow of water in order to correct for the apparent increase in level. Of course, on increasing steam demand we need an increased water flow. The pressure in the drum will ultimately be restored, e.g. by a pressure controller on the steam header increasing the boiler duty, and the level controller will ultimately increase the water flow. However, for the level controller to be stable, the initial process behaviour means that it will have to act far more slowly than the tight controller defined in Equation (4.16). The converse of swell is known as *shrink*, where a decrease in steam demand causes the drum pressure to temporarily rise and the water level to apparently decrease.

This problem may be solved by using a dp-type level instrument that effectively measures the mass of liquid in the drum rather than its volume. Since the effect is caused by a reduction in the fluid density,

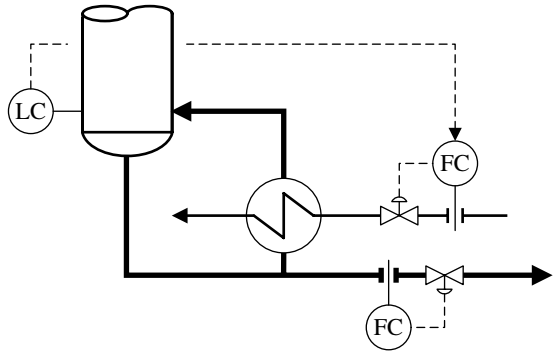

Figure 4.26 *Control of column level by manipulation of reboiler duty*

rather than an increase in its inventory, an instrument measuring the head of liquid will respond correctly. However, some of the increase in level may be due to bubbles expanding in the tubes supplying the drum forcing additional water into the drum. Further local legislation may dictate, for safety reasons, that actual liquid level must be measured and used for control. Under these circumstances the problem can be alleviated by applying a correction term to the level measurement.

$$L_{corrected} = L_{measured} + K\left(P_{measured} - P_{normal}\right) \tag{4.53}$$

The term K is determined empirically from process data and has the effect of increasing the level measurement transmitted to the controller when the measured pressure increases above the normal operating pressure.

The terms *shrink* and *swell* are also used also to describe the inverse response that can arise with level control. The boiler feed water ideally is heated in the economiser to the boiling point of water at drum pressure. However, this is often not achieved so that when the cooler water enters the drum it will cause a drop in temperature, thus causing bubbles to collapse and the water level to drop. While controllers can generally be tuned to handle inverse response, they have to be tuned to act more slowly to avoid instability.

Three-element level control (Figure 4.27) is a technique that introduces a feedforward element into the controller. It includes a measurement of the steam flow leaving the drum. Any change in this flow is immediately passed to the water flow so as to maintain the mass balance. This largely meets the objective of tight level control. The level controller is retained as a feedback controller to compensate for any flow measurement errors and to allow the operator to change the SP if required. It may now be tuned to act relatively slowly.

The feedforward and feedback signals are traditionally combined by using a *bias* algorithm which simply adds the signals. Details of this and the alternative *ratio* algorithm are included in Chapter 6. The bias algorithm requires that the two flow measurements are in the same engineering units. Where water flow is measured in m^3/hr and steam flow in te/hr, this is already taken care of. If inconsistent units are used, then a scaling factor will be required.

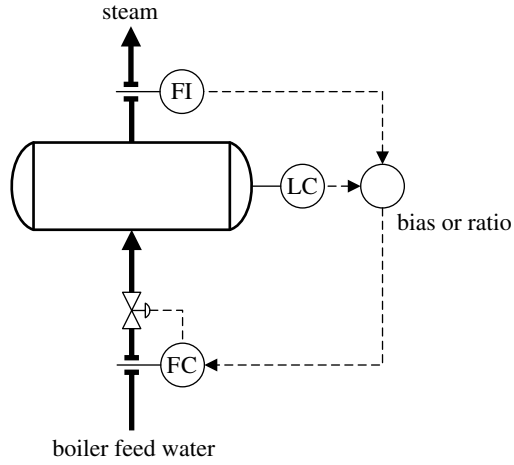

Figure 4.27 *Three-element level control*

One approach, effectively the same as the use of a bias algorithm, is to calculate the difference between the steam and water flows – again in consistent units. This is then the PV of a flow difference controller which manipulates the control valve on the water supply. Theoretically its SP should be zero but is manipulated by the cascaded level controller.

The ratio algorithm multiplies the two signals, effectively keeping the water flow in an adjustable proportion to the steam flow – where this proportion can be in any units. The disadvantage of applying ratio control is that the process gain varies as the steam flow varies. We know that, for constant steam flow, the rate of change of level is related to the flow of water.

$$\frac{dL}{dt} \propto F_{water} \tag{4.54}$$

By definition

$$R = \frac{F_{water}}{F_{steam}} \tag{4.55}$$

And so

$$\frac{dL}{dt} \propto F_{steam} R \tag{4.56}$$

Comparing Equation (4.56) with Equation (4.48), where the PV is now L and the MV is R, gives

$$K_p \propto F_{steam} \tag{4.57}$$

To maintain a constant loop gain, K_c must be kept inversely proportional to K_p. It would appear, therefore, that we would have to adjust the controller gain if the steam flow changes. However, since the level controller is less critical with the feedforward scheme in place, the use of the range of the steam meter instead of F_{steam} will result in conservative tuning that will be stable over the whole operating range.

Care should be taken if ratio feedforward is implemented as an enhancement to an existing level controller. Since the level controller will now be manipulating the ratio target, rather than the flow controller SP, its controller gain may need adjusting. The controller, working in dimensionless form, will generate a change in output (ΔM) which is converted to flow change in engineering units (ΔF) using the ranges of the flow and level controllers, i.e.

$$\Delta F = \Delta M \times \frac{FC_{range}}{LC_{range}} \tag{4.58}$$

With the ratio in place the range of the ratio algorithm replaces that of the flow controller and the change in ratio target is converted to change in water flow by multiplying it by the measured steam flow (F_{steam}), i.e.

$$\Delta F = \Delta M \times \frac{R_{range}}{LC_{range}} \times F_{steam} \tag{4.59}$$

This will change the effective controller gain. To compensate for this, the existing controller gain should be multiplied by

$$\frac{FC_{range}}{R_{range} \times F_{steam}} \tag{4.60}$$

As above, the range of the steam controller should be used in place of the actual steam flow. If the instrumentation is well engineered then this range will be similar to that of the water. The range of the ratio is chosen on configuration. Since the actual ratio will change little then it should be possible to choose a range so that Equation (4.60) generates a correction factor close to unity – thus avoiding any adjustment of tuning. This is of particular benefit if the operator is permitted to selectively disable the feedforward part of the scheme, e.g. because of a problem with the steam flow instrument, since it would avoid the need to switch between two values for controller gain.

5

Signal Conditioning

Signal conditioning is manipulation of the input measurement to (or output signal from) a controller. A mathematical function is applied in order to improve controller performance. It may be required to compensate for nonlinear behaviour or to remove measurement noise. Other process parameters may also be incorporated into the PV to improve the accuracy of control.

5.1 Instrument Linearisation

The most frequent application of signal conditioning is linearisation. Many of the common functions may not be obvious to the control engineer since they are often built into the DCS or transmitter as standard features. For example, where c_d is the *discharge coefficient*, d the orifice diameter, dp the pressure drop across the orifice and ρ the fluid density, the flow (F) through an orifice flow meter is given by

$$F = c_d \frac{\pi d^2}{4} \sqrt{\frac{dp}{\rho}} \tag{5.1}$$

The flow can therefore be measured by measuring dp but, to ensure that there is a linear relationship between this and the flow, the square root of dp is used. This is known as *square root extraction* and is usually an option within the DCS, or it might be performed by the field transmitter. Its effect is illustrated in Figure 5.1.

There is a similar need for linearisation of temperature measurements by *thermocouples*. These rely on the *Seebeck Effect* which generates a voltage at the junction of two dissimilar metals. The metals are often alloys. For example, the *J type* thermocouple is a junction of iron and *constantan* – an alloy of copper (55%) and nickel (45%). The *K type* uses *chromel* (90% Ni, 10% Cr) and *alumel* (95% Ni, 2% Mn, 2% Al, 1% Si). Their calibration curves are shown in Figure 5.2. Although over much of the range the relationship between temperature and voltage is linear, this is not the case for temperatures below 0°C. Standard conversion tables are published for each thermocouple type and these are usually incorporated into the DCS or transmitter.

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

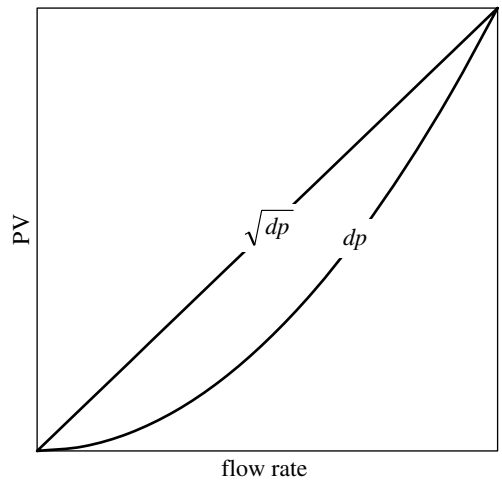

Figure 5.1 *Square root extraction for flow meter*

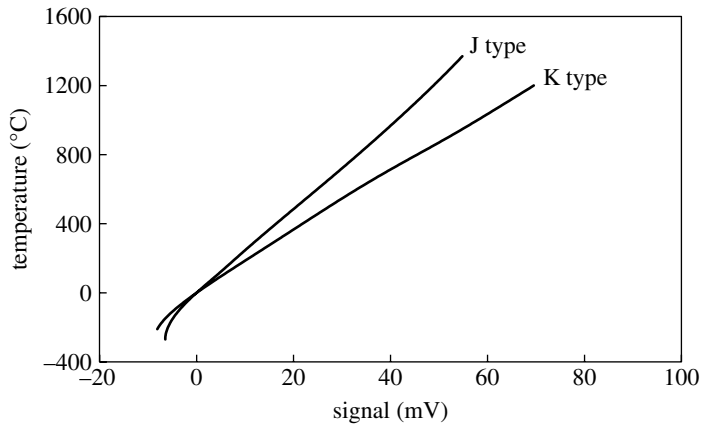

Figure 5.2 *Thermocouple calibration*

Resistance Temperature Detectors (RTD) use a different linearisation function. The *Callendar-van Dusen Equation* relates resistance (R) to temperature (T) according to

$$R = R_0\left[1 + AT + BT^2 + C\left(T - 100\right)T^3\right] \quad \text{for} \quad -200^oC \le T \le 0^oC \tag{5.2}$$

$$R = R_0\left[1 + AT + BT^2\right] \quad \text{for} \quad 0^oC \le T \le 850^oC \tag{5.3}$$

The coefficients A, B and C depend on the metal used. The most common is commercial grade platinum with a nominal resistance of 0.385 Ω/°C, in which case:

$$A = 3.9083 \times 10^{-3} \qquad B = -5.775 \times 10^{-7} \qquad C = -4.183 \times 10^{-12} \tag{5.4}$$

The resistance at 0°C (R_0) is determined by the thickness of the wire. For the most common type, described as Pt100, R_0 is 100Ω. Thicker wire, Pt10, is often used for very high temperatures. Nickel can also be used. Figure 5.3 shows some typical calibrations.

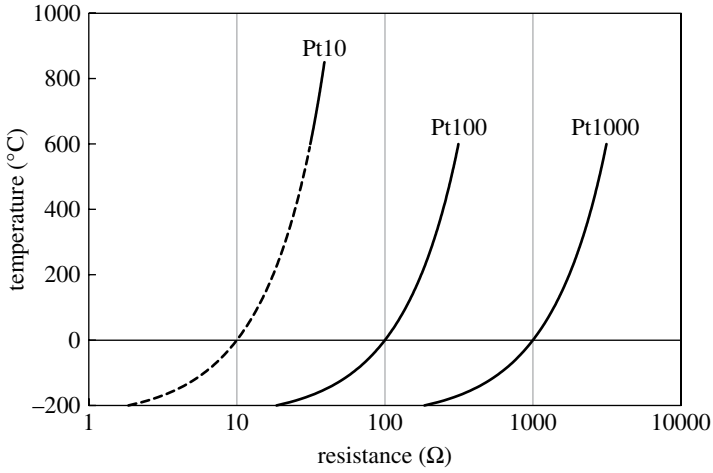

Figure 5.3 RTD calibration

5.2 Process Linearisation

Signal conditioning can also be applied to compensate for nonlinear process behaviour. For example, in Chapter 4 we covered the linearisation of a poorly engineered level gauge so that it truly represented the % utilisation of the vessel's working volume and so its rate of change would be linearly related to the manipulated flow.

Figure 5.4 shows a typical problem caused by nonlinearity. Three equal steps were made to the SP but the response becomes increasingly oscillatory – because the process gain becomes larger as the process moves into the operating region where the PV (and MV) are higher. In fact, we can estimate the change in process gain from the steady state conditions. The first increase of 10% in the SP required the MV to increase from 10 to 21%, giving a process gain of around 0.9. The second step required the MV to increase by a further 5% to 26%, giving a process gain of around 2.0. The final step resulted in the MV increasing from 26% to 30%, giving a process gain of around 2.5. Good control, with approximately a threefold

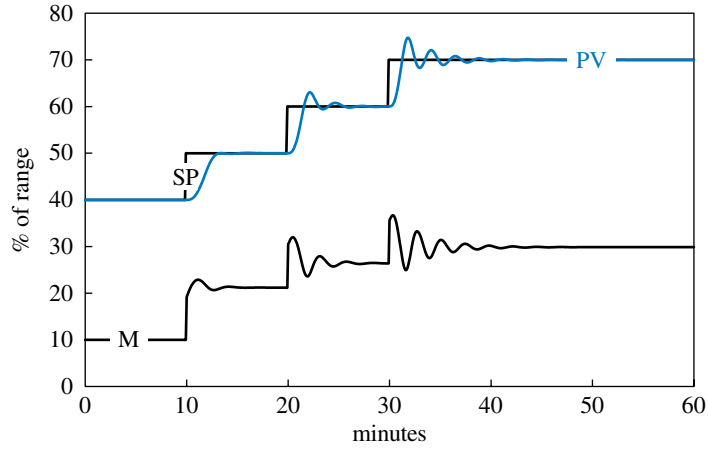

Figure 5.4 Effect of nonlinearity

change in process gain, would be impossible to achieve without some form of linearisation. In fact, the process gain varies more than this analysis would indicate. The gain at a PV of 40% will be less than that determined over the range of 40–50%. Similarly the gain at a PV of 70% will be greater than 2.5.

Our aim is to develop a linearisation function that can be applied to the PV to generate *PV**, such that

$$PV^* = K_p MV + bias \tag{5.5}$$

We require K_p to be as constant as possible. We are free to choose target values for both K_p and *bias*; for simplicity here we choose 1 and 0 respectively. We are also free to choose the form of the linearisation function. If a polynomial, it must be at least second order.

$$PV^* = a_0 + a_1 PV + a_2 PV^2 = MV \tag{5.6}$$

The coefficients (a) are obtained by regression analysis between *PV* and *MV*. In our example they are

$$a_0 = -69.7 \qquad a_1 = 2.77 \qquad a_2 = -0.0193 \tag{5.7}$$

Figure 5.5 shows both PV (as the black points) and *PV** (as the coloured points). As the MV increases the process gain of the linearised PV varies between 1.25 and 0.75. While just outside the maximum variation of ±20% that we typically require, it would allow effective controller tuning. Figure 5.6 shows the effect of the linearisation function with the controller retuned to take it into account. While the effect of the nonlinearity is still apparent, particularly the slower response to the last SP change, control over the whole range is substantially improved. Should further improvement be required, a higher order polynomial could be used. But a better approach, similar to that described in the next section, would be to use process modelling to develop a function based on anticipated process behaviour.

Often the first indication of nonlinearity is the failure to reliably identify process dynamics from plant testing. However, there is no need to repeat the tests with the linearisation in place; it can be retrospectively applied to the data already collected. Care should be taken with the choice of linearising function to ensure it behaves well over the whole operating range. Carelessly applied, the function could cause the process gain to change wildly – possibly reversing sign. The new measurement (PV^*) will probably have

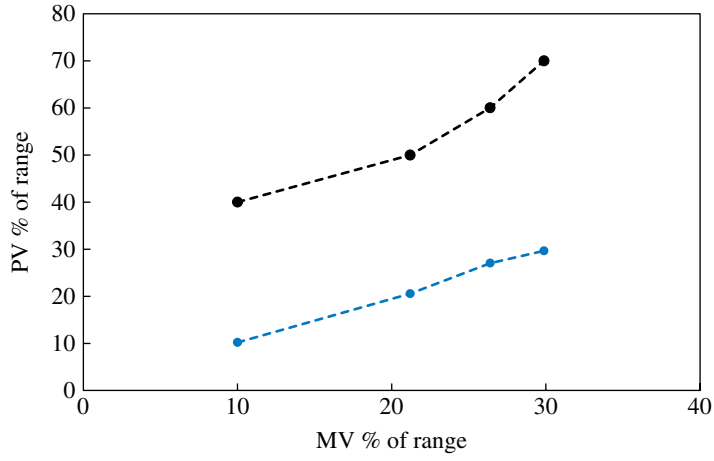

Figure 5.5 *PV linearisation*

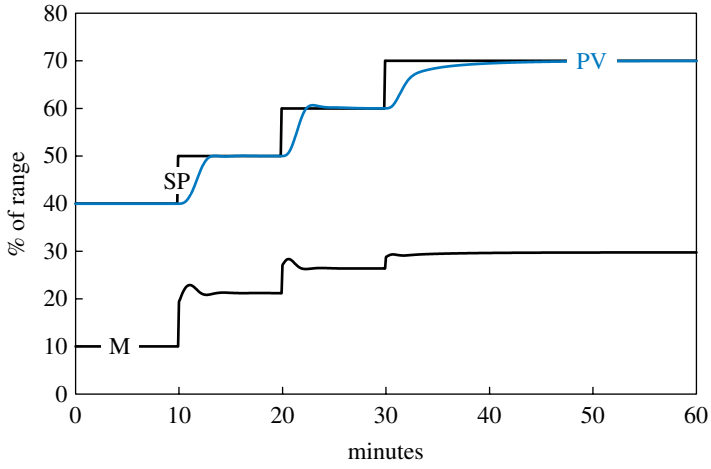

Figure 5.6 *Impact of linearising PV*

no engineering meaning so, rather than display the value to the operator, the same linearisation function should be applied to the SP.

5.3 Control of pH

Perhaps the most challenging nonlinear control problem is that of pH. Figure 5.7 shows (as the solid lines) the curves for a strong base, of pH 13, being titrated against a strong acid, of pH 2, and against another of pH 1.5. This illustrates two problems. As shown by the solid line in Figure 5.8, with an acid pH of 2, the process gain (in engineering units) varies from a value of 0.00042 to 197. This variation is by a factor over 300,000 times larger than can handled by a linear controller. Secondly, if the flow of base is correct for neutrality (i.e. the pH is 7), a change of 0.5 in the pH of the acid would mean we move vertically from one curve in Figure 5.8 (where K_p is around 200) to the other (where K_p drops to around 0.002). Again, a linear controller could not handle such a load change.

It might appear that the nonlinearity might be characterised by dividing the titration curve into several sections that can be treated as linear. For example, the section between a pH of 4 and 10 would appear to be a straight line. However, as Figure 5.9 shows, when zooming in on this section of the curve, it is actually very nonlinear. The process gain changes by a factor of around 50. And again it is tempting to assume that between a pH of 6 and 8, the line is straight. But, as Figure 5.10 shows, this also is not the case; the process gain varies by a factor of around 6.

To derive a linearising function we first need to understand the process in more detail. By definition, pH is the negative logarithm of the concentration (in kg-ions/m^3) of hydrogen ions, i.e.

$$pH = -\log_{10}\left[H^+\right] \qquad \text{or} \qquad \left[H^+\right] = 10^{-pH} \tag{5.8}$$

Pure water ionises:

$$H_2O \leftrightarrow H^+ + OH^- \tag{5.9}$$

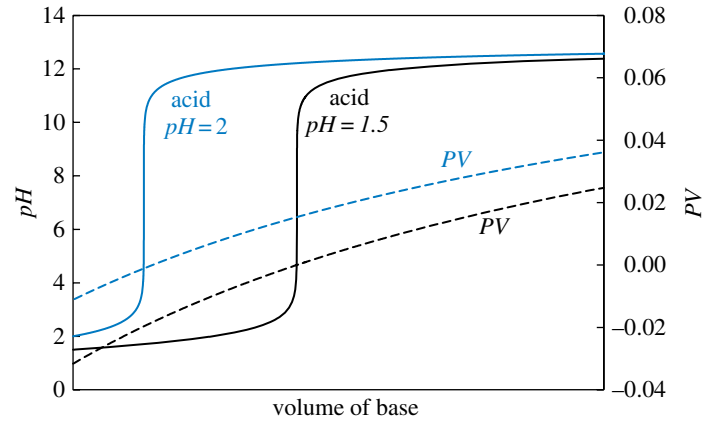

Figure 5.7 *Nonlinearity of pH*

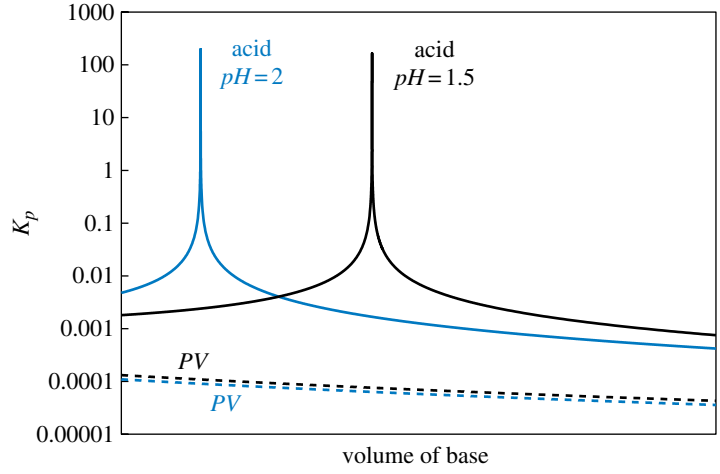

Figure 5.8 *Variation of process gain with pH*

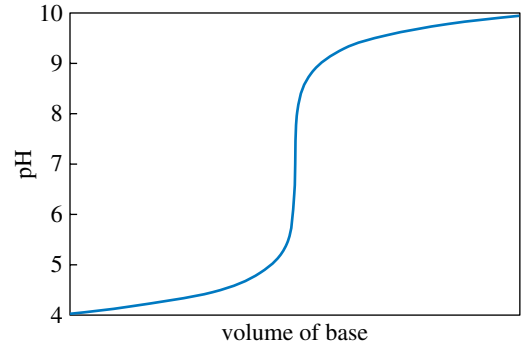

Figure 5.9 *Nonlinearity of pH when close to neutrality*

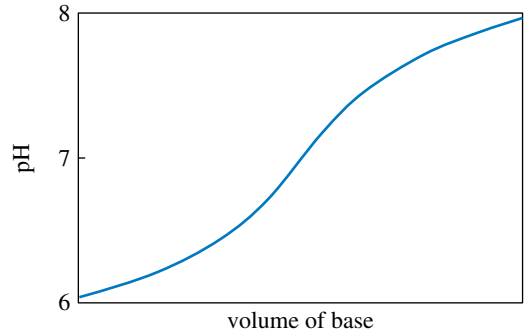

Figure 5.10 *Non-linearity of pH when even closer to neutrality*

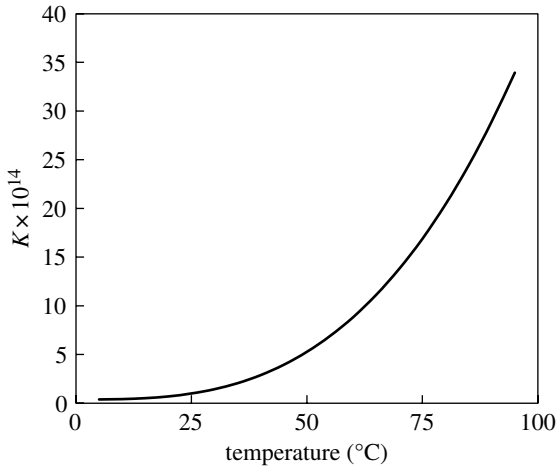

Figure 5.11 *Effect of temperature on dissociation of water*

where the equilibrium constant *(K)* is defined as

$$K = \frac{[H^+][OH^-]}{[H_2O]} \tag{5.10}$$

Water is only weakly ionised and so $[H_2O]$ is effectively 1. Figure 5.11 shows how K varies with temperature. At 25°C, the ionisation is such that the equilibrium constant for water (K_w) is 10^{-14} and so, from Equations (5.8) and (5.10)

$$[OH^-] = 10^{pH-14} \tag{5.11}$$

If K_a is the equilibrium constant for the ionisation of an acid HA then

$$[HA] = \frac{[H^+][A^-]}{K_a} = 10^{-pH}\frac{[A^-]}{K_a} \tag{5.12}$$

Similarly, if K_b is the equilibrium constant for the ionisation of a base BOH then

$$[BOH] = \frac{[B^+][OH^-]}{K_b} = 10^{pH-14}\frac{[B^+]}{K_b} \tag{5.13}$$

Total acid concentration is given by

$$[HA] + [H^+] = 10^{-pH}\left(\frac{[A^-]}{K_a} + 1\right) \tag{5.14}$$

Total base concentration is given by

$$[BOH] + [OH^-] = 10^{pH-14}\left(\frac{[B^+]}{K_b} + 1\right) \tag{5.15}$$

We define a PV as the difference between the base and acid concentrations, i.e.

$$PV = 10^{pH-14}\left(\frac{[B^+]}{K_b} + 1\right) - 10^{-pH}\left(\frac{[A^-]}{K_a} + 1\right) \tag{5.16}$$

This would have a value of zero at neutrality. For a mixture of strong acid and strong base:

$$K_a \to \infty \qquad \text{and} \qquad K_b \to \infty \tag{5.17}$$

And so

$$PV = 10^{pH-14} - 10^{-pH} \tag{5.18}$$

The dashed lines in Figures 5.7 and 5.8 show the result of applying this formula. While this does not give a perfectly linear relationship, it is a considerable improvement. The much more modest change in process gain should not present a tuning problem. Further the process gain will change little as acid strength changes. Figure 5.12 shows the closed loop response for an optimally tuned pH controller based on this

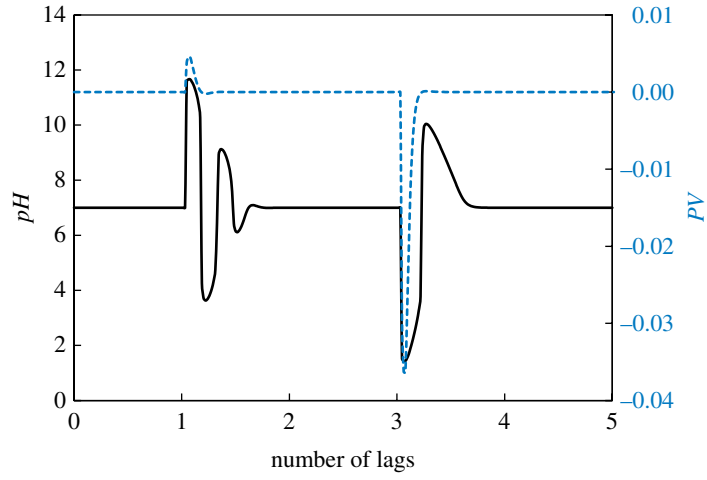

Figure 5.12 *Typical closed loop response for pH control*

linearisation technique. The first disturbance caused the pH to increase rapidly, from its SP of 7, to about 12. The second caused a reduction to about 1. Provided the deadtime is close to zero, the controller can be tuned to return the pH to SP within one process lag. The oscillatory nature of the response, not obvious in the linearised PV, is unavoidable if this fast return to SP is required. The shape of the oscillation will be different depending on the direction of the disturbance.

Linearisation becomes more important in the control of pH in a batch process since it is important that the SP is not overshot. Once added, the neutralising reagent cannot be removed to bring the pH back to SP.

In the case where either the acid or the base is weak (or if there are other salts present) *buffering* will occur. This reduces the effect that changing either has on pH. PV linearisation must therefore be based on the titration curve determined in the laboratory. Characterisation might be implemented as a look-up table or by regressing coefficients added to Equation (5.18).

Because of the logarithmic relationship, correction of a disturbance caused by a change of 1 in the inlet pH typically requires a tenfold change in the manipulated flow. Disturbances up to 3 would therefore require flow control with a 1000:1 turndown – beyond the range of a single controller. One solution, particularly if the variation in pH is larger than this, is a multi-stage process where the pH is changed in two or three stages – each with its own controller. For more modest variation, two flow controllers with different ranges can be manipulated by a single pH controller – as described in Chapter 8.

5.4 Constraint Conditioning

Signal conditioning can be used to extend the apparent range of a measurement. It is common in constraint control applications to use the output (M) of a PID controller as an indication of valve position. This is a measure of how close the process is to a hydraulic limit. The problem is that, if the constraint is being violated, the controller output will be 100% – no matter how bad the violation.

If the controller is saturated then the PV will not usually be at SP. The size of the error (E) gives an indication of the severity of the problem. We can incorporate this into the measurement of the constraint (PV).

$$PV = M + K.E \tag{5.19}$$

This is illustrated as an example in Figure 5.13. Imagine that we wish to maximise the feed rate to our case study heater (Figure 2.10) and that the constraint in doing so is a hydraulic limit on the fuel. As feed rate is increased the temperature controller will take corrective action and increase the signal (M) to open the fuel valve – usually via a cascade to a flow controller. This is a *hard constraint*, i.e. it can only be approached from one side – it is mechanically impossible for the valve to exceed an opening of more than 100%. If the heater was operating close to this limit and there was a process disturbance, for example a drop in the heater inlet temperature, the temperature controller would increase its signal to the valve, potentially taking it to the 100% limit.

The problem is that, if we wish to alleviate the constraint, the 100% indication does not tell us how far the constraint has been violated. The heater could be operating exactly at the true maximum feed rate, or could be well beyond it. It is for this reason that the process operator will typically set a maximum limit of around 90% – effectively turning the hard constraint into a soft one. This of course reduces the achievable feed rate. However, if the unit is operating well beyond the real 100% constraint, the heater outlet temperature will be below its SP and we can incorporate the temperature controller error (E) as a measure of the severity of the violation. In this example K would be set to the negative reciprocal of the process gain between the outlet temperature and the signal to the valve. In doing so the PV, as defined in Equation (5.19) can now exceed 100% and its relationship to feed rate will have the same process gain as it does so. While it would still not be advisable for the process operator to now set the limit at 100%, the conditioning will improve confidence in the controller and permit at least some increase.

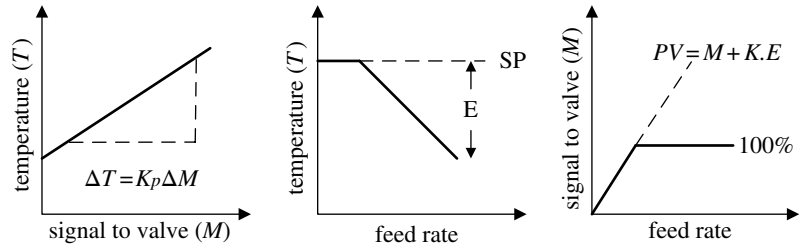

Figure 5.13 *Constraint conditioning (hard constraint)*

A similar approach can be applied to the measurement of flue gas oxygen. If the air-to-fuel ratio falls below the stoichiometric requirement then the oxygen analyser will indicate zero – no matter how bad the problem. In Chapter 10 we show how incorporating a measurement of carbon monoxide (CO) can apparently extend the range into negative values of oxygen content.

There are occasions where a nonlinear response is preferred. We may want a controller to respond more quickly if the PV moves away from SP in a particular direction. For example, we can adjust K in Equation (5.19) so that violation of a constraint is dealt with more quickly than it is approached. Similarly, even if the measurement stays within range, we may be more concerned about a high PV than a low one. We could again use the error to condition the measurement.

$$PV = PV_{measured} + K\left(PV_{measured} - SP\right) \tag{5.20}$$

In this example, K is set to zero if the measured PV is less than the SP, otherwise it is set to a value typically around 0.3 and so effectively increases K_c by 30%. Care should be taken in introducing such nonlinearities so that control remains stable when operating in the region where the conditioning is active.

Again, as an example, let us imagine that the constraint on increasing feed rate to our case study heater is now a limit on maximum burner pressure. Unlike the fuel valve position, this is a *soft constraint*. Although violation is undesirable, it is physically possible. Burner pressure will continue to rise as feed rate is increased. However, high burner pressure can extinguish the flame and would be considered hazardous. So, any violation should be dealt with more urgently than exploiting spare capacity. Figure 5.14 illustrates how Equation (5.20) would be applied, increasing the apparent severity of the violation.

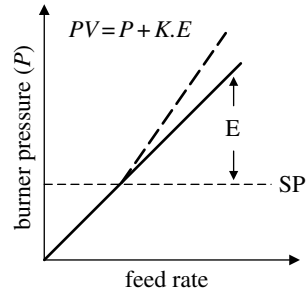

Figure 5.14 *Constraint conditioning (soft constraint)*

5.5 Pressure Compensation of Distillation Tray Temperature

Many process measurements are sensitive to pressure changes. By incorporating the pressure measurement into the PV we can ensure that the controller takes the correct action. We have already covered one example of this as Equation (4.53) – a means of reducing the effect of swell and shrink in steam drums by conditioning the level measurement to reduce its sensitivity to pressure.

We can adopt a similar approach to tray temperature controllers on distillation columns. They provide some control of product composition because this correlates with the bubble point of the liquid. However, changing pressure changes this relationship. Figure 5.15 shows the effect pressure has on bubble point, in this case water, but all liquids show similar behaviour.

If a distillation tray temperature controller keeps the temperature constant as the pressure changes, the composition will move away from target. We can resolve this by using the pressure to condition the temperature measurement. The subject of *pressure compensated temperatures* is covered in full in Chapter 12.

5.6 Compensation of Gas Flow Measurement

Gas flow measurements, from a meter that relies on creating a pressure drop (dp), are sensitive to pressure, temperature and the molecular weight of the gas. The instrument range, configured in the DCS, was determined assuming a calibration pressure (P_{cal}), temperature (T_{cal}) and molecular weight (MW_{cal}). If the current conditions (P, T and MW) are different from these then we must apply a correction to the measured flow ($F_{measured}$) to obtain the true flow (F_{true}). The form of correction depends on the units of measure. Equation (5.21) should only be applied when the flow is recorded in volumetric units at standard conditions, e.g. nm^3/hr or SCFM (standard cubic feet per minute). Pressure and temperature should be on an absolute basis.

$$F_{true} = F_{measured}\sqrt{\frac{MW_{cal}}{MW} \times \frac{P}{P_{cal}} \times \frac{T_{cal}}{T}} \tag{5.21}$$

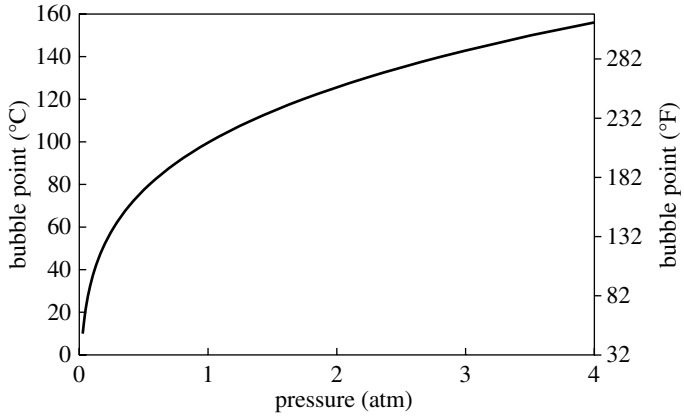

Figure 5.15 *Effect of pressure on bubble point*

If the flow measurement is in actual volumetric units, i.e. reported at actual (rather than standard) pressure and temperature, then the formula becomes

$$F_{true} = F_{measured}\sqrt{\frac{MW_{cal}}{MW} \times \frac{P_{cal}}{P} \times \frac{T}{T_{cal}}} \tag{5.22}$$

And if the flow measurement is on a weight basis then

$$F_{true} = F_{measured}\sqrt{\frac{MW}{MW_{cal}} \times \frac{P}{P_{cal}} \times \frac{T_{cal}}{T}} \tag{5.23}$$

Meters to which Equations (5.21) to (5.23) apply include orifice plates, pitot tubes, venturis and annubars. However, the formulae should be applied with care. In Chapter 10 we show how their application to gaseous fuels can worsen problems with combustion control.

Similarly applying them to gas mixtures, where the aim is to maintain the flow of a single component as composition changes, also requires special consideration. For example, if we wished to control the flow of hydrogen supplied as a mixture with other gases then we can infer, from MW, the mole fraction of hydrogen (x). Knowing the molecular weight of the gas mixed with the hydrogen (MW_{other}) and that of hydrogen itself ($MW_{hydrogen}$) gives

$$x = \frac{MW_{other} - MW}{MW_{other} - MW_{hydrogen}} \tag{5.24}$$

So the flow of hydrogen, in standard volumetric units, is given by

$$F_{true} = F_{measured}\sqrt{\frac{MW_{cal}}{MW} \times \frac{P}{P_{cal}} \times \frac{T_{cal}}{T}} \times \frac{MW_{other} - MW}{MW_{other} - MW_{hydrogen}} \tag{5.25}$$

To demonstrate the importance of this modification, let us assume that under calibration conditions the gas contained 70 vol% hydrogen ($MW_{hydrogen} = 2$) and 30 vol% methane ($MW_{other} = 16$); therefore MW_{cal} is 6.2. If the hydrogen content increases to 80 vol% then MW will be 4.8. Assuming that pressure and temperature remain at calibration conditions, then applying Equation (5.21) shows that the total gas flow is about 14% higher than that measured. Applying Equation (5.25) shows that the increase in the flow of hydrogen is much higher at 30%.

There are metering techniques that measure directly the actual volumetric flow of gas (F_{actual}). These include vortex shedding, turbine, ultrasonic and magnetic types. To convert their measurement to volume at standard conditions we correct for pressure and temperature.

$$F_{true} = F_{actual}\sqrt{\frac{P}{P_{standard}} \times \frac{T_{standard}}{T}} \tag{5.26}$$

There are also meters, including coriolis and thermal types, which measure directly the mass flow of the gas (F_{mass}). To convert their measurement to standard volumetric units we need only allow for changes in molecular weight. Density at standard conditions ($\rho_{standard}$) can be derived by applying the Ideal Gas Law to give:

$$\rho_{standard} = \frac{P_{standard}.MW}{R.T_{standard}} \tag{5.27}$$

R is the *Universal Gas Constant*. It has a value of 8.314 kJ/kg-mole/K (1.9859 BTU/lb-mole/R). Care must be taken in choosing consistent units of measure for pressure. For example, in SI units, these must be kPa. If working in imperial units, then the units of pressure are likely to be psi. A multiplier of 0.185 should be included for the resulting density to be in lb/ft^3. Standard volumetric flow is then given by

$$F_{true} = F_{mass} \times \frac{R.T_{standard}}{P_{standard}.MW} \tag{5.28}$$

Some models of coriolis meters also provide density as an additional measurement. This can be used to determine the actual volumetric flow. But because it is determined at stream, rather than standard, conditions it cannot be used as a measure of molecular weight unless stream pressure and temperature are also measured.

Whether the correction term for temperature should be included should be given careful consideration. If, for example, the calibration temperature is 50°C (around 120°F) and the actual temperature varies by 10% then, because we convert both the calibration and actual temperatures to an absolute basis and then take square root of the ratio, the error introduced to the volumetric flow is less than 1%. This is probably within the measurement repeatability. Temperature correction is only worthwhile therefore if the operating temperature is very high, or if the change can be very large.

A similar argument applies to pressure compensation if the operating gauge pressure is close to zero. For example, a 10% change in a pressure of 0.3 barg (around 4 psig), when converted to absolute pressure, causes a flow measurement error of about 1%.

Unnecessarily including additional measurements increases the probability that an instrument failure will affect the process. It also adds complexity, making technical support more difficult – particularly if the designer of the scheme has moved on.

The pressure, temperature and (particularly) the molecular weight measurements should be validated and, if found faulty, the last good measurement should be used in the flow compensation. This *graceful degradation* causes no disturbance to the process and requires no action by the operator. However, if the compensated flow is used in a controller, care must be taken in recommissioning the repaired instrument. Its measurement is likely to be different from the value currently being used; so, unless the controller is forced to reinitialise, the change in value will cause a process disturbance.

If the compensated flow is to be used as the PV of a controller, then the formulae defined in this section can be applied as written. Some plant owners, however, take the view that the raw uncompensated measurement should still be used by the controller. Instead of compensating the measurement, the SP can be compensated by multiplying it by the reciprocal of the applicable formula.

5.7 Filtering

Another common form of signal conditioning is filtering – used to reduce measurement noise. Noise may be genuine in that the instrument is faithfully reproducing rapid fluctuations in the measurement. Examples include measuring the level of a turbulent liquid or the flow of a mixed phase fluid. Noise may be introduced mechanically by vibration or electrically though interference. While filtering may reduce the problem, it is unlikely to remove it completely and it will distort the base signal. Whatever the cause, efforts should be made to eliminate the noise at source. The use of baffles or stilling wells around the level sensor can prevent turbulence affecting the measurement. Ensuring that flows are measured where liquid is below its bubble point will avoid flashing across the orifice plate. Placing transmitters away from vibrating equipment and having signal cables properly screened and not routed close to large electrical equipment will avoid induced noise.

Filtering will change the apparent process dynamics, usually in a way detrimental to controller performance. This is often explained in text books, using Figure 5.16, as a phase lag. If a sinusoidal signal is injected into a conventional DCS filter then the output will be reduced in amplitude and shifted in time. This is not particularly helpful in the process industry, where the engineer rarely comes across sinusoidal signals. Perhaps a more pragmatic approach is to consider the noisy measurement trended in Figure 5.17. The challenge is to remove the noise from the underlying base signal, without distorting it. This would appear straightforward enough; most could add the underlying base signal to the trend shown. But doing so involves looking back in time. If the second half of the trend had yet to be drawn, it would be quite difficult to decide whether the downward movement is a genuine reduction or just another noise spike. It is not until more information is provided that the distinction will be clear. Filters have the same problem; it is not until well after the base signal has changed that the filter can recognise it. There will therefore be a delay of some sort, before the change is passed to the controller.

While a noisy measurement may not look good when trended, this no reason to add a filter. The criterion on which the decision should be made is how much noise is passed through the controller to the final actuator, e.g. the control valve. If there is a danger of mechanical damage then a filter may offer the only practical solution. In particular, filtering should be considered if derivative action is justified since this would otherwise greatly amplify the noise.

The problem is that no filter is perfect. While all filters can be tuned to suit the process, there will always be a compromise between noise reduction and base signal distortion. Whether a filter is effective depends

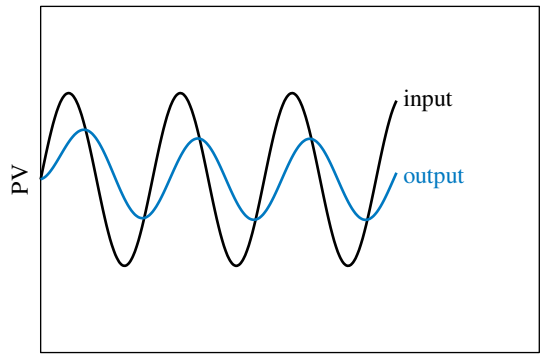

Figure 5.16 *Filter phase lag*

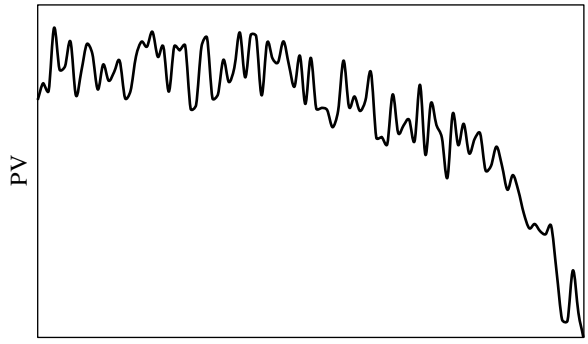

Figure 5.17 *Typical noisy measurement*

on the relative impact these two problems have on the controller. Filter lag may be of little concern if the process already has a very large lag. Noise reduction may not be critical if the controller gain is small and there is no need for derivative action.

5.8 Exponential Filter

DCS have generally standardised on the *first order exponential filter*. In other applications, such as *statistical process control (SPC)*, it can also be described as the *geometric mean* or the *exponentially weighted moving average (EWMA) filter*. This introduces an engineer-configurable lag (with time constant τ_f) on the PV. It is implemented as

$$Y_n = P.Y_{n-1} + (1-P)X_n \tag{5.29}$$

Y_n is the current output from the filter, Y_{n-1} the previous output and X_n the current input. P is a tuning parameter set by the engineer in the range 0 to 1. If set to 0, the current output will be equal to the current input and no filtering takes place. If set to 1, the current output will be equal to the previous output and any change in measurement is ignored. Some systems permit any value within this range; others limit P to predefined values such as 0, 0.5, 0.75 or 0.85. Other systems accept the time constant (τ_f) where this is related to P by

$$P = e^{-ts/\tau_f} \qquad \text{or} \qquad \tau_f = \frac{ts}{-\ln(P)} \tag{5.30}$$

Some texts define P as

$$P = \frac{\tau_f - ts}{\tau_f} \qquad \text{or} \qquad \tau_f = \frac{ts}{1-P} \tag{5.31}$$

This is based on the first order Taylor approximation

$$e^{-ts/\tau_f} = 1 - \frac{ts}{\tau_f} \tag{5.32}$$

Another similar definition of P is

$$P = \frac{\tau_f}{\tau_f + ts} \qquad \text{or} \qquad \tau_f = \frac{P.ts}{1-P} \tag{5.33}$$

This is based on the reciprocal first order Taylor approximation

$$e^{-ts/\tau_f} = \frac{1}{e^{ts/\tau_f}} = \frac{1}{1+\frac{ts}{\tau_f}} \tag{5.34}$$

And others will define P as

$$P = \frac{2\tau_f - ts}{2\tau_f + ts} \qquad \text{or} \qquad \tau_f = \frac{(1+P)ts}{2(1-P)} \tag{5.35}$$

This is based on the *first order Padé approximation*, more details of which can be found in Chapter 15.

$$e^{-ts/\tau_f} = \frac{2 - \frac{ts}{\tau_f}}{2 + \frac{ts}{\tau_f}} \tag{5.36}$$

Remembering that ts is likely to be measured in seconds and τ_f in minutes, ts will generally be very much smaller than τ_f. Higher order terms in the Taylor and Padé approximations will then rapidly approach zero. The performance of a filter based on any of the three approximations will be indistinguishable from the exact version unless τ_f is very small – in which case one would question whether the filter is necessary.

An example of the effectiveness of the filter is illustrated in Figure 5.18. While the actual level of noise reduction is dependent on the type and frequency of the noise, it shows that it is approximately linear with P. However, the impact on process dynamics is highly nonlinear. Figure 5.19 shows that the lag introduced by the filter increases sharply as the value of P exceeds 0.9. Should this level of filtering be required then the additional lag may cause a problem. If it exceeds around 20% of the process lag then, even if the controller is re-tuned to accommodate it, the degradation in performance will be noticeable. Under these circumstances it would be better to adopt a different technique, such as one of those covered later in this chapter.

While selection of the value of P is usually determined by trial and error, some have published techniques that make an initial estimate – provided the frequency of the noise is known. We will show in Chapter 15 that the *amplitude ratio* or *attenuation* (K_f) of a sinusoidal signal of frequency f passing through a first order lag of τ_f is given by

$$K_f = \frac{1}{\sqrt{\left(2\pi f.\tau_f\right)^2 + 1}} \tag{5.37}$$

While little used by most process control engineers, plotting K_f against f produces half of a *Bode diagram* showing the *frequency response*. This is shown as black lines in Figure 5.20. As might be expected, it

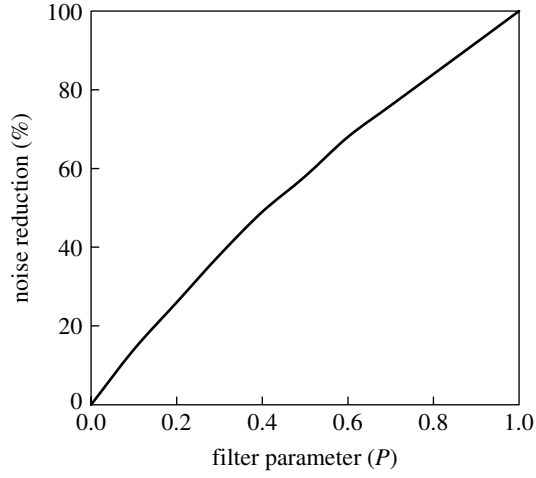

Figure 5.18 *Performance of the exponential filter*

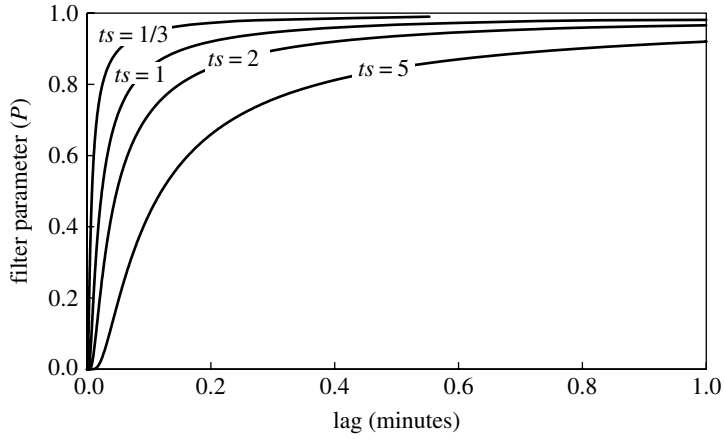

Figure 5.19 *Impact of controller scan interval on filter lag*

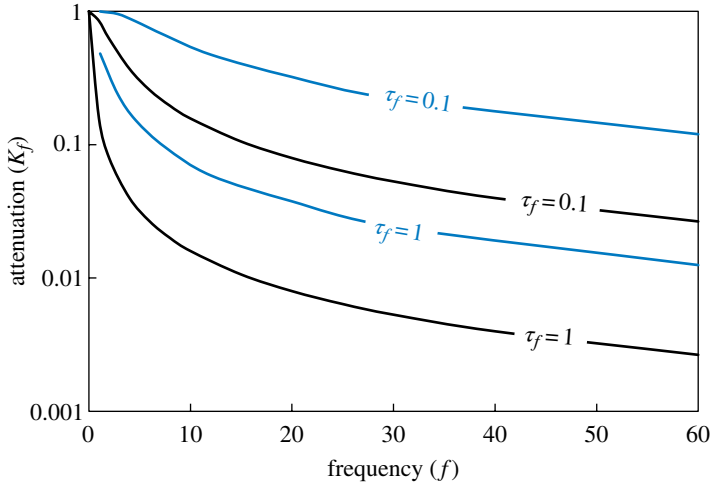

Figure 5.20 *Bode diagram*

shows that higher frequency noise requires a smaller filter lag which will then distort the base signal less. Outside of the process industry, attenuation is often expressed in decibels, defined as $20.\log(K_f)$. Although not of importance here, the other half of the Bode diagram is a plot of the *phase shift* angle (ϕ) caused by passing through the lag.

$$\phi = -\tan^{-1}\left(2\pi f.\tau_f\right) \tag{5.38}$$

Rearranging Equation (5.37)

$$\tau_f = \frac{\sqrt{\frac{1}{\left(K_f\right)^2} - 1}}{2\pi f} \tag{5.39}$$

For example, if we require a 90% reduction in noise amplitude, we would choose a value of 0.1 for K_f. At such low values Equation (5.39) can be approximated to

$$\tau_f = \frac{1}{2\pi f.K_f} \tag{5.40}$$

Noise frequency can be estimated by counting the number of peaks that occur within a known interval – one minute for example. However, Equation (5.40) should only be used to provide an initial estimate for τ_f. It is unlikely that the noise will be sinusoidal and may not have a constant frequency or amplitude. Further the frequency may be reduced by *aliasing* which occurs if the filter scan frequency is less than half the noise frequency. It also assumes analog control. The filter lag required to achieve the required noise reduction may therefore be substantially higher. To illustrate this, the coloured lines in Figure 5.20 show more realistic curves based on a real example. It is also difficult to predict the attenuation required since the noise transmitted to the final control element will also depend on controller scan interval and tuning.

An advantage of choosing this filter over others described later in this chapter is that a historised filtered measurement can be unfiltered. Rearranging Equation (5.29) gives

$$X_n = \frac{Y_n - PY_{n-1}}{(1-P)} \tag{5.41}$$

This might be applied if the engineer suspects that excessive filtering has been applied and wants to explore the impact of removing or reducing the filter. If, during step-testing, data were collected with the filter in place, applying Equation (5.41) would permit the dynamics to be determined for the unfiltered measurement. The controller could then be tuned to take advantage of the improved dynamics arising from removal of a filter.

Care needs to be taken when calculating P from τ_f, or vice versa, to work in consistent units of time – since scan intervals are usually quoted in seconds, while lags are generally in minutes. As we saw in Figure 5.19, the relationship between P and τ_f depends on the controller scan interval (*ts* seconds). While it is unusual to change the scan interval of the DCS, it is common for controllers to be moved from an older to a newer system that may have a different scanning frequency. The filter will then perform differently either in terms of noise reduction or the effect it has on the apparent process dynamics. Either way, the performance of the controller may degrade.

To understand this, consider a DCS where the filter is defined by setting τ_f rather than P. If we require a 90% reduction in noise then, using the example of Figure 5.18, we would choose a value of 0.87 for P. Assuming this was implemented in the original DCS with a scan interval of 0.33 second then, from Equation (5.30), this is equivalent to choosing a value for τ_f of 0.04 minutes. If the DCS is replaced with one that has a scan interval of 2 seconds then retaining the same value for τ_f would, from Equation (5.30), be equivalent to reducing P to 0.43. From Figure 5.18 this would result in the noise reduction falling to about 50%. Figure 5.21 shows the impact of switching to other common scan intervals. To restore P to 0.87 to achieve the same level of noise reduction would require, from Equation (5.30), τ_f to be increased to 0.24 minutes. Conversely, in a system where the filter is defined by setting P, the change in scan interval will not affect noise reduction but the filter lag will still increase from 0.04 to 0.24 minutes. If significant compared to the process lag, this might require the controller to be re-tuned. The situation can be further complicated by the effect that scan interval has on how noise is passed through the controller to the actuator, as described by Figure 3.65. The engineer should not underestimate the re-retuning of filters and controllers that can be required if a control system upgrade results in a plant-wide change in scan interval.

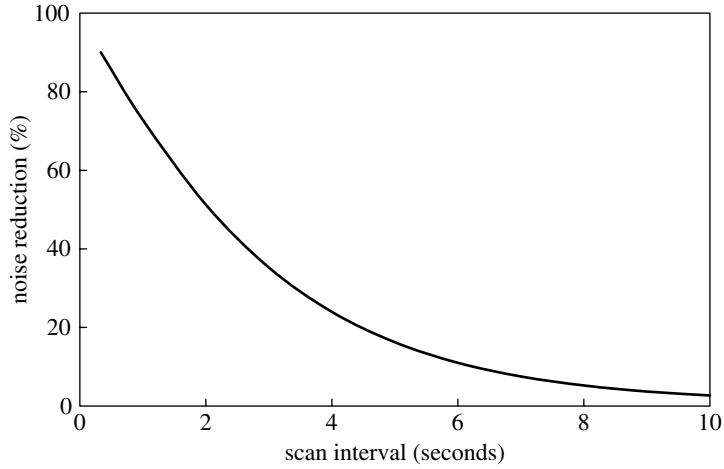

Figure 5.21 *Impact of controller scan interval on noise reduction*

5.9 Nonlinear Exponential Filter

It is possible to modify the first order exponential filter, described by Equation (5.29), to make it nonlinear. Instead of the parameter P being set by the engineer, it is changed automatically according to the formulae

$$|X_n - Y_{n-1}| > R \qquad \text{then} \qquad P = 0 \tag{5.42}$$

$$|X_n - Y_{n-1}| \leq R \qquad \text{then} \qquad P = 1 - \frac{|X_n - Y_{n-1}|}{R} \tag{5.43}$$

The engineer selects the value for R. If the difference between the current input and the last output is greater than this value then the change passes through unfiltered. Changes less than R are filtered depending on their size. Very small changes are heavily filtered while larger changes are filtered less. The objective of this enhancement is to reduce the lag caused by filtering. Figure 5.22 shows its effectiveness at noise reduction for different types of noise. The filter works well if the noise amplitude is predictable. By setting R to a value slightly higher than this amplitude, real changes in the base signal will be little affected by the filter. This would be the situation, for example, if the noise in a flow measurement is caused by flashing across an orifice plate. However, the filter offers little advantage if the noise is less predictable or 'spiky' – such as level measurement of a boiling liquid. Such noise is often described as *Gaussian*, reflecting its statistical distribution. To prevent the spikes passing through the filter, the value of R has to be set so large that the filter behaves much like the unmodified exponential filter. Its performance on our example process is shown in Figure 5.23.

5.10 Moving Average Filter

Simple averaging can be used as a filter. The filter can be represented as

$$Y_n = \frac{\sum_{r=1}^{N} X_{N-r+1}}{N} \tag{5.44}$$

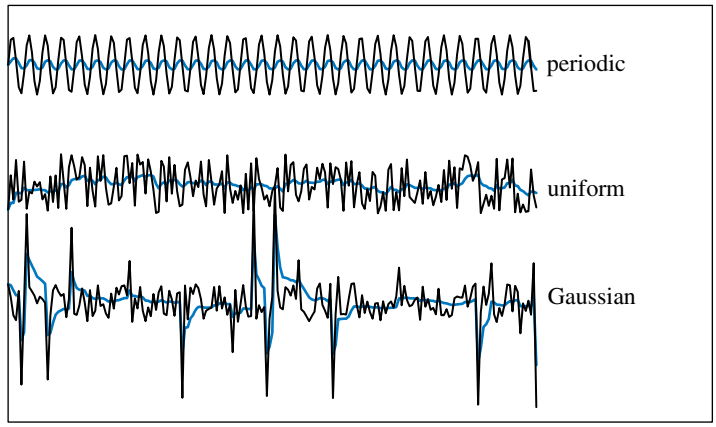

Figure 5.22 *Impact of noise type on performance of nonlinear exponential filter*

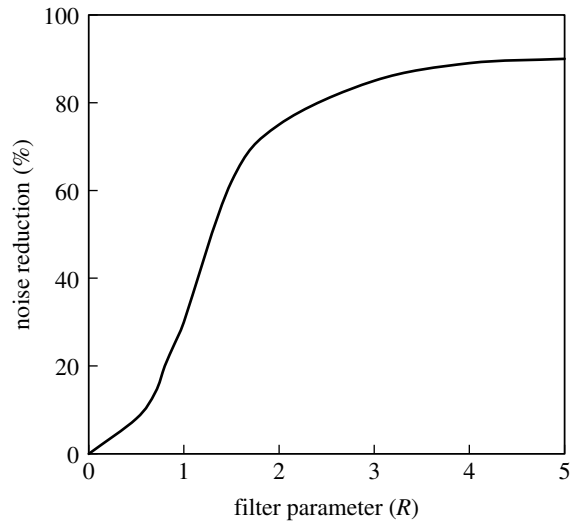

Figure 5.23 *Performance of the nonlinear exponential filter*

N is the number of historical values included in the average and is the tuning constant defined by the engineer. The filter can also be written as

$$Y_n = B_1 X_N + B_2 X_{N-1} \ldots . + B_r X_{N-r+1} \ldots . + B_N X_1 \tag{5.45}$$

The filter should have a gain of 1 and so

$$\sum_{i=1}^{N} B_i = 1 \tag{5.46}$$

And because it is a simple linear average

$$B_1 = B_2 \ldots . = B_{N-r+1} \ldots . = B_N = \frac{1}{N} \tag{5.47}$$

This filter offers little advantage over the standard exponential filter. Rather than a lag, it introduces a ramp function where the duration of the ramp is given by *N.ts*. This would be visible if the input were a step change. But when superimposed on the process lag the result, with *N* adjusted to give equivalent noise reduction, is virtually indistinguishable from the exponential filter – as shown in Figure 5.24. Indeed it can be shown that it closely approximates to an exponential filter with a lag of 0.534*N.ts*. Given that the moving average filter is unlikely to be standard feature of the DCS and that its impact on process dynamics is not quite so easy to predict, it offers little advantage.

5.11 Least Squares Filter

We can choose different coefficients for *B* in Equation (5.45) provided they sum to 1. Indeed, the exponential filter described by Equation (5.29) can be written as

$$Y_n = (1-P)X_n + P(1-P)X_{n-1} + P^2(1-P)X_{n-2} \ldots \tag{5.48}$$

To show that these coefficients sum to 1, we can apply the formula for the sum of a geometric progression (where *r* is the *common ratio* and *a* the *scale factor*).

$$\sum_{i=m}^{n} ar^i = ar^m + ar^{m+1} + ar^{m+2} \cdots + ar^n = \frac{a\left(r^m - r^{n+1}\right)}{1-r} \tag{5.49}$$

$$\sum_{i=0}^{n} (1-P)P^i = \frac{(1-P)\left(P^0 - P^n\right)}{1-P} \tag{5.50}$$

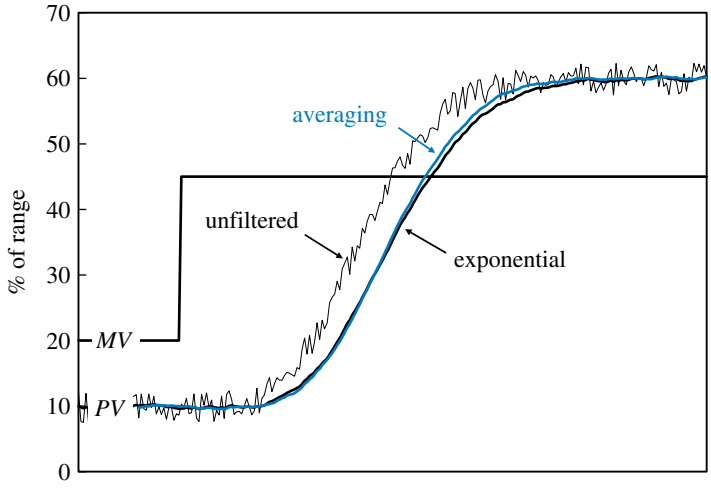

Figure 5.24 *Comparison of averaging versus exponential filter*

Since P is between 0 and 1, P^n will approach zero as n approaches infinity and so Equation (5.50) evaluates to 1. Unfortunately the series is infinite. While it would be possible to choose a value for n where the sum of coefficients (S) is close enough to 1, the length of the series is likely still to prove impractical. This is given by

$$n = \frac{\log(1-S)}{\log(P)} \tag{5.51}$$

While a filter with P set to 0.7 needs only 13 historical values for the sum of the coefficients to exceed 0.99, one with P set to 0.98 will need 228. However, an approach which uses this technique to advantage is the *least squares filter*. It gets its name from the least squares regression technique used to fit a line to a set of points plotted on a XY (scatter) chart. Its principle, as shown in Figure 5.25, is to fit a straight line to the last N points. The end of this line is Y_n.

We can show that the filter is of the form of Equation (5.45). The development of the formula to estimate the coefficients $B_1, B_2 \ldots B_N$ is quite complex. However, the end result is simple to apply. Indeed, the reader could now skip to Equation (5.65), if happy to just to accept the result.

Figure 5.25 shows the last N values of a process measurement (Y). We define the time axis as r, where $r = 1$ for the most recent measurement and $r = N$ for the oldest. The filter is based on predicting the value of Y based on the equation of the line of best fit, where m is the slope of the line and c the intercept on the Y axis, i.e.

$$\hat{Y} = mr + c \tag{5.52}$$

The equation of the line is developed to minimise the sum of the squares between the predicted value of Y and the actual value, i.e.

$$\sum_{i=1}^{N} E_i^2 = \sum_{i=1}^{N} \left(\hat{Y}_i - Y_i\right)^2 = \sum_{i=1}^{N} \left(mr_i + c - Y_i\right)^2 \tag{5.53}$$

Partially differentiating with respect to each of m and c, and setting the derivative to 0 will identify the best choice of these values, i.e.

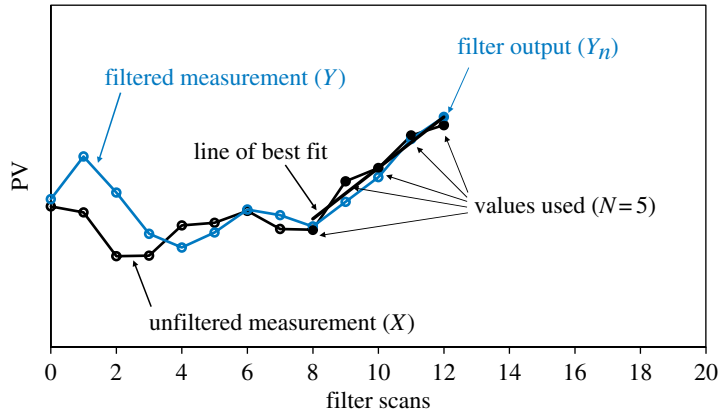

Figure 5.25 *Principle of least squares filter*

$$\frac{\partial \sum_{i=1}^{N} E_i^2}{\partial m} = \sum_{i=1}^{N} \left(2r_i^2 m + 2rc - 2rY_i\right) = 0 \tag{5.54}$$

$$\therefore \quad m\sum_{i=1}^{N} r_i^2 + c\sum_{i=1}^{N} r_i - \sum_{i=1}^{N} r_i Y_i = 0 \tag{5.55}$$

$$\frac{\partial \sum_{i=1}^{N} E_i^2}{\partial c} = \sum_{i=1}^{N} \left(2r_i m + 2c - 2Y_i\right) = 0 \tag{5.56}$$

$$\therefore \quad m\sum_{i=1}^{N} r_i + Nc - \sum_{i=1}^{N} Y_i = 0 \tag{5.57}$$

Solving Equations (5.55) and (5.57) gives

$$m = \frac{\sum_{i=1}^{N} r_i \sum_{i=1}^{N} Y_i - N\sum_{i=1}^{N} r_i Y_i}{\left(\sum_{i=1}^{N} r_i\right)^2 - N\sum_{i=1}^{N} r_i^2} \tag{5.58}$$

$$c = \frac{\sum_{i=1}^{N} r_i \sum_{i=1}^{N} r_i Y_i - \sum_{i=1}^{N} r_i^2 \sum_{i=1}^{N} Y_i}{\left(\sum_{i=1}^{N} r_i\right)^2 - N\sum_{i=1}^{N} r_i^2} \tag{5.59}$$

We wish to predict the current value of Y, i.e. when $r = 1$ and so

$$\hat{Y} = m + c \tag{5.60}$$

And so

$$\hat{Y} = \frac{\sum_{i=1}^{N} r_i \sum_{i=1}^{N} Y_i - N\sum_{i=1}^{N} r_i Y_i + \sum_{i=1}^{N} r_i \sum_{i=1}^{N} r_i Y_i - \sum_{i=1}^{N} r_i^2 \sum_{i=1}^{N} Y_i}{\left(\sum_{i=1}^{N} r_i\right)^2 - N\sum_{i=1}^{N} r_i^2} \tag{5.61}$$

This is a linear function of previous values of Y and so can be written in the form of Equation (5.45). To determine the coefficients (B), we use the formula for the sum of a series of consecutive integers.

$$\sum_{i=1}^{N} r_i = \frac{N\left(N+1\right)}{2} \tag{5.62}$$

We also use the sum of a series of squares of consecutive integers

$$\sum_{i=1}^{N} r_i^2 = \frac{N(2N+1)(N+1)}{6} \tag{5.63}$$

Substituting these into Equation (5.61) we can determine the value of each coefficient, so

$$B_r = \frac{\frac{N(N+1)}{2} - Nr + \frac{N(N+1)}{2} r - \frac{N(2N+1)(N+1)}{6}}{\left(\frac{N(N+1)}{2}\right)^2 - N\frac{N(2N+1)(N+1)}{6}} \tag{5.64}$$

This simplifies to

$$B_r = \frac{4(N+1) - 6r}{N(N+1)} \tag{5.65}$$

The disadvantage of applying any filter other than the standard first order exponential type is that it will require custom coding in the DCS. However, in the case of the least squares type, the calculation is quite simple and the coefficients (B) can be determined outside of the DCS and stored as constants. As an illustration, Equation (5.65) has been used to generate the coefficients in Table 5.1 for values of N up to 8. It can be seen that setting N to 1 effectively removes the filter, but so does setting N to 2. This is because fitting a straight line to two points will always pass through both points and so Y_n will be the same as X_n.

We will show shortly that the value required for N will be greater than that required in a simple averaging filter, leading the user to suspect that that filter lag will be larger. However, the advantage of the least squares filter is that, because it uses the trend of the unfiltered measurement to predict the filtered value, it has no lag. Indeed, its predictive nature introduces *lead* which partially counteracts the process lag. Figure 5.26 shows, depending on the value chosen for N, how the filter distorts a noise-free step change. It causes an overshoot of around 30% and will clearly change the apparent process dynamics. Figure 5.27 shows that low values of N cause substantially less overshoot. However, as Figure 5.28 shows, a larger value of N is required to achieve the same level of noise reduction as the moving average filter. It is unlikely that values of N less than 15 will normally be of benefit.

Table 5.1 *Coefficients for least squares filter*

N	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8
1	1.000							
2	1.000	0.000						
3	0.833	0.333	–0.167					
4	0.700	0.400	0.100	–0.200				
5	0.600	0.400	0.200	0.000	–0.200			
6	0.524	0.381	0.238	0.095	–0.048	–0.190		
7	0.464	0.357	0.250	0.143	0.036	–0.071	–0.179	
8	0.417	0.333	0.250	0.167	0.083	0.000	-0.083	-0.167

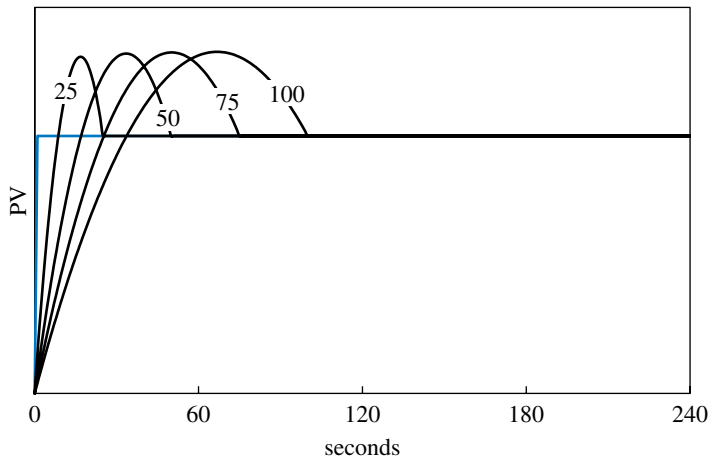

Figure 5.26 *Impact of least squares filter on step change*

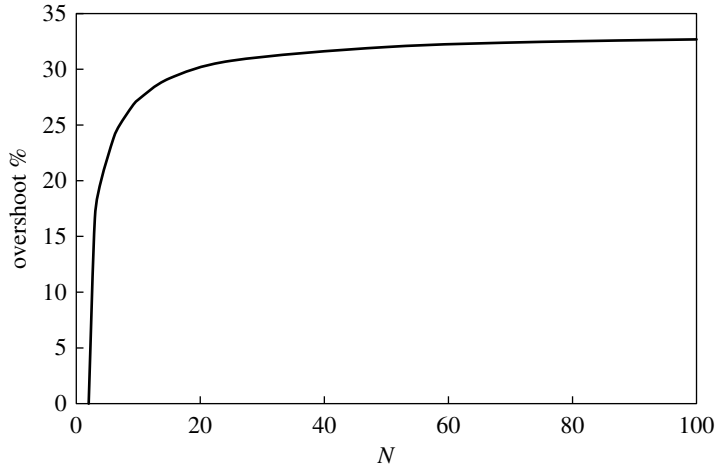

Figure 5.27 *PV overshoot caused by least squares filter*

Figure 5.29 shows the same filters as Figure 5.26, only this time applied to the more usual lagged measurement. In this case the lag is 30 seconds, with a filter scan interval of one second. With N set to 50, it is likely that the process can still be modelled as first order and so the tuning method suggested in Chapter 3 would be reliable. However, a significantly higher value of N might present a problem. The change in dynamics is not only a function of N but also of the process lag and the filter scan interval. Performance should therefore be evaluated carefully before controller tuning. However in the right circumstances, as shown in Figure 5.30, it will significantly outperform the exponential filter. Both filters have been tuned to give the same level of noise reduction. The least squares filter not only outperforms the exponential filter in that it tracks the base signal more closely but it overtakes the real measurement, so reducing the process lag.

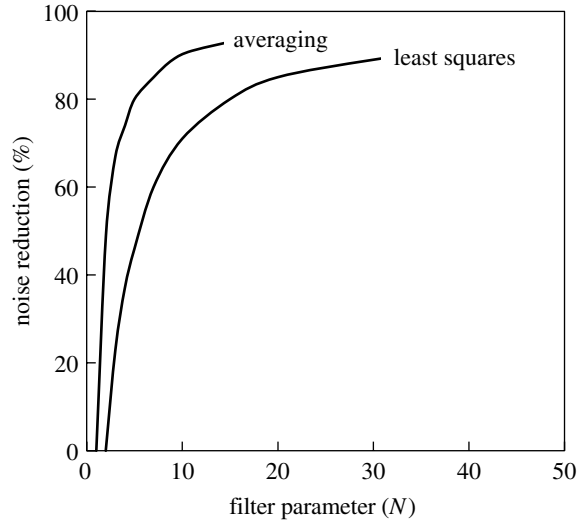

Figure 5.28 *Performance of averaging and least squares filters*

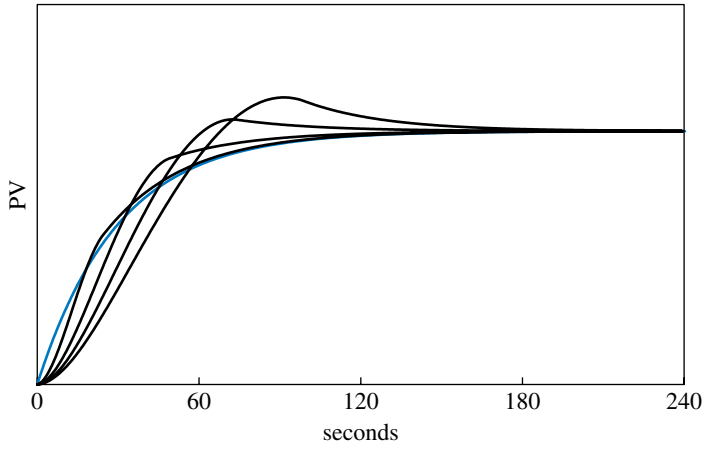

Figure 5.29 *Impact of least squares filter on process dynamics*

We will see in Chapter 9 that it can be useful to know the slope (m) of the line of best fit. Rearranging Equation (5.65) gives

$$r = \frac{\left(4 - NB_r\right)\left(N + 1\right)}{6} \tag{5.66}$$

Substituting Equations (5.62), (5.63) and (5.66) into Equation (5.58) gives

$$m = \frac{2\left(Y_n - \bar{X}\right)}{1 - N} \tag{5.67}$$

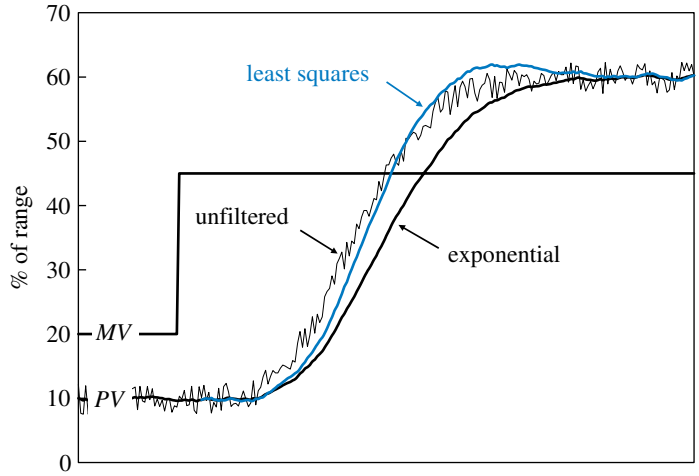

Figure 5.30 *Comparison of least squares versus exponential filter*

Remember that, because r varies with time from N to 1, m will be opposite in sign to the true rate of change of X. Further, the slope will have units of change in X per scan interval (ts). Dividing the slope by ts is necessary to give the true rate of change. We can see that the slope is derived from the current filtered measurement and the mean of the last N unfiltered measurements. This mean is also the output of the averaging filter described in the previous section. By substituting Equations (5.65) and (5.47) into Equation (5.67), the rate of change can therefore also be calculated as

$$\frac{dX}{dt} = \frac{B_1 X_N + B_2 X_{N-1} \ldots . + B_r X_{N-r+1} \ldots . + B_N X_1}{ts} \tag{5.68}$$

$$B_r = \frac{-2}{1-N}\left(\frac{4N-6r+4}{N(N+1)} - \frac{1}{N}\right) = \frac{6(N-2r+1)}{N(N^2-1)} \tag{5.69}$$

5.12 Tuning the Filter

In addition to correctly selecting the filter type, the challenge is to tune the filter to give the best overall controller performance. The controller gain generally has to be reduced to accommodate the change in process dynamics resulting from the addition of the filter. There will therefore be a trade-off between noise reduction and speed of controller response. As we saw in Chapter 3, Li suggests (Equation 3.143) modifying the penalty function (e.g. ITAE) used for controller tuning to take account of total valve travel. The latter is a measure of filter effectiveness and its inclusion permits the combined optimisation of filter tuning and controller tuning. Instead of total valve travel, some suggest using the total number of changes in the direction of valve travel. In practice this is affected little by filtering which largely reduces the amplitude of the noise but has little impact on frequency.

Figure 5.31 shows the effect of changing the lag (τ_f) of the conventional first order exponential filter. In each case the controller tuning has been optimised with the filter in place. The choice of weighting factor significantly changes the optimum filter lag – increasing it from 5% to 25% of the process lag as the valve

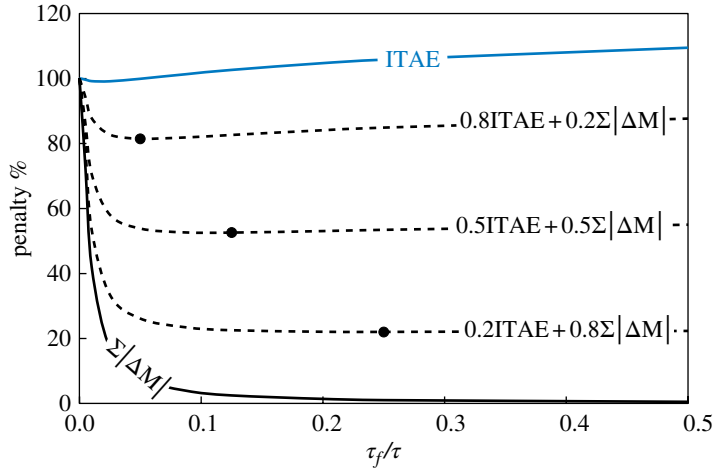

Figure 5.31 *Optimisation of filter tuning*

weighting factor is increased from 0.2 to 0.8. However, no matter which weighting is chosen, the penalty function increases very little as the filter is increased above 5% of the process lag – suggesting that, in this case, this would be the best choice.

5.13 Control Valve Characterisation

Nonlinear valves are an example of *output conditioning*. It is often the case that we do not want the valve position to move linearly with controller output. The most common type of nonlinear control valve is the *equal percentage* type, often used when controlling the flow of fluid from a centrifugal pump or turbo-compressor – as shown in Figure 5.32.

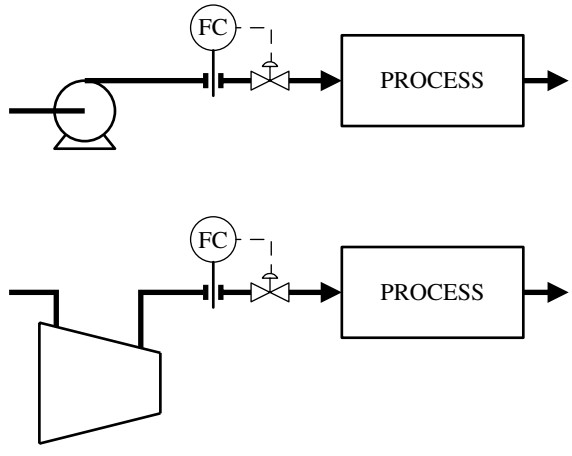

Figure 5.32 *Control of pump and compressor discharge flow*

Figure 5.33 shows how pressure varies with flow. Assuming that the pressure at the process exit is constant, e.g. because the product is routed to storage, then the process inlet pressure will increase with the square of the flow. Also shown is the pump curve. With no control valve in place, the pump discharge pressure must be the same as the process inlet pressure and the flow is set by where the pump curve and the process curve intersect. But we want to control the flow at a desired value. To do so, the drop in pressure (Δp) across the control valve must be equal to the pump discharge pressure less the process inlet pressure.

Figure 5.34 shows the effect that varying the flow has on the pressure drop across the valve. Ideally we want a linear relationship between flow and valve position, as would be the result if the pressure drop

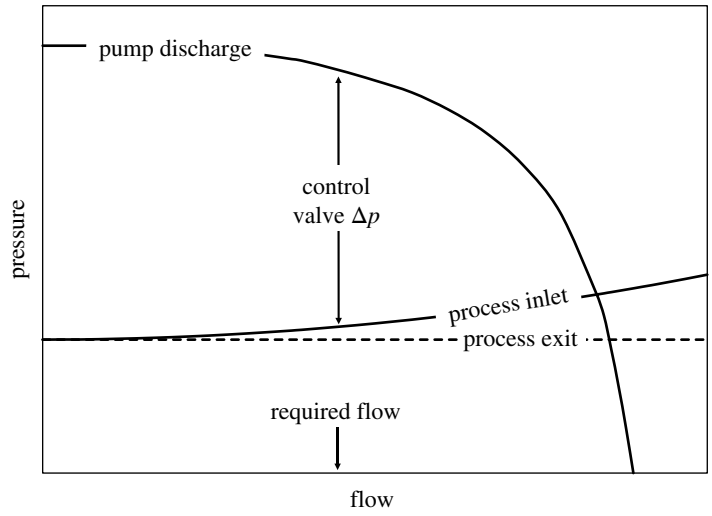

Figure 5.33 *System pressures*

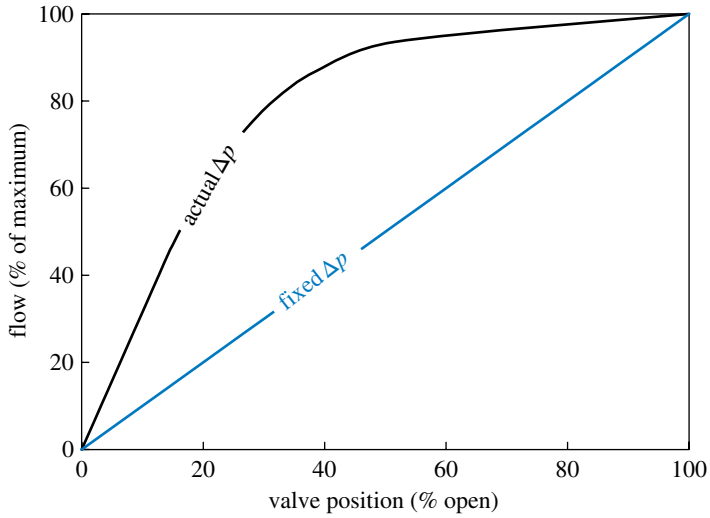

Figure 5.34 *Relationship between flow and valve position*

across the valve were constant. Figure 5.35 shows ways in which the controller output can be conditioned to compensate for the nonlinearities introduced by the process and pump curves. Some DCS permit the definition of a look-up table, sometimes called a *function generator*. For example, the DCS would interpolate between the five specified points – effectively joining them with straight lines to closely match the required curve.

5.14 Equal Percentage Valve

An alternative approach, also shown in Figure 5.35, is the use of an equal percentage valve. This may be manufactured to behave in a specified nonlinear behaviour or it may be a conventional linear valve fitted with a positioner in which the engineer can define the valve characterisation. The definition of 'equal percentage' is that the same change in valve position will give the same percentage change in flow. If, as a result of the valve position being changed from V_1 to V_2, the flow changes from F_1 to F_2 then for a linear valve

$$(F_2 - F_1) \propto (V_2 - V_1) \tag{5.70}$$

In the case of an equal percentage valve the relationship between flow and valve position becomes

$$\frac{(F_2 - F_1)}{F_1} \propto (V_2 - V_1) \tag{5.71}$$

This illustrated in Figure 5.36; the change in valve position required to increase the flow from 40 to 60 (i.e. a 50% increase) is the same as that required to increase the flow from 10 to 15 (i.e. also a 50% increase). Equation (5.71) cannot be applied when F_1 is zero, so a small compromise has to be made to ensure the

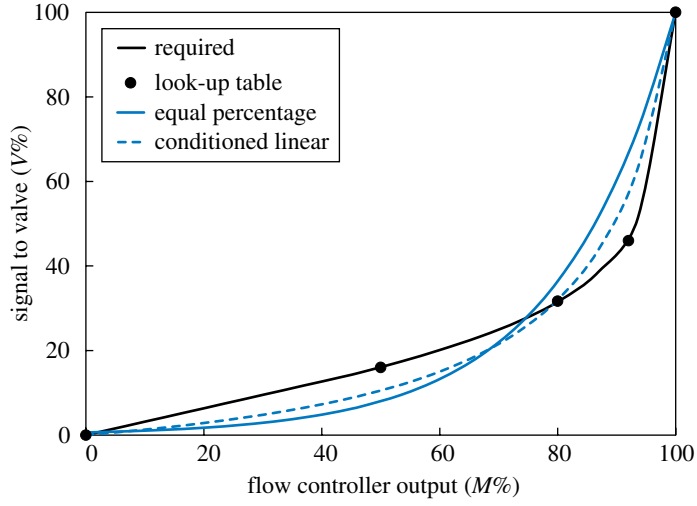

Figure 5.35 *Valve characterisation*

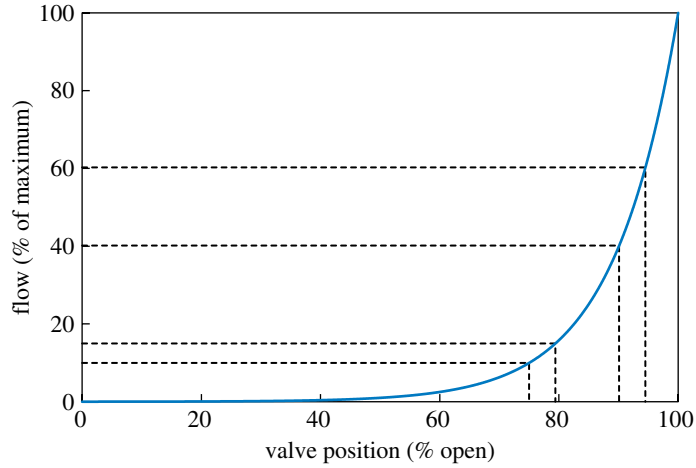

Figure 5.36 *Equal percentage valve*

flow through the valve is zero when the valve is fully shut but, in general, if F is the flow and V the % valve position, then

$$\frac{\Delta F}{F} \propto \Delta V \tag{5.72}$$

And so, by introducing the *valve constant (k)*

$$\frac{dF}{F} = kdV \tag{5.73}$$

Integrating gives

$$\ln\left(F\right) = kV + A \qquad \text{or} \qquad F = e^{(kV+A)} \tag{5.74}$$

The constant A can be eliminated by defining F_{max} as the flow with the valve 100% open, i.e.

$$F_{max} = e^{(100k+A)} \tag{5.75}$$

And so

$$\frac{F}{F_{max}} = \frac{e^{(kV+A)}}{e^{(100k+A)}} = e^{k(V-100)} \tag{5.76}$$

An alternative form of the transformation is based on the *rangeability (R)*.

$$\frac{F}{F_{max}} = R^{\frac{V}{100}-1} \qquad \text{where} \qquad R = e^{100k} \quad \text{and} \quad k = \frac{\ln\left(R\right)}{100} \tag{5.77}$$

The process gain (between flow and valve position) of a linear valve is constant. For an equal percentage valve, rearranging Equation (5.73) shows that its process gain is proportional to the flow.

$$K_p = \frac{dF}{dV} = kF \tag{5.78}$$

If ρ is the liquid density then control valves are characterised by the *flow coefficient* (C_v) defined as

$$C_v = F\sqrt{\frac{\rho}{\Delta p}} \tag{5.79}$$

C_v strictly is defined as the volumetric flow in USGPM (US gallons per minute) of water at a temperature of 60°F with a pressure drop across the valve of 1 psi. There is a SI equivalent (K_v) defined as the volume flow in m^3/hr of water at a temperature between 5° and 40°C with a pressure drop across the valve of 1 bar. The two definitions are related by

$$K_v = 0.865C_v \tag{5.80}$$

For a linear valve C_v is constant, but for an equal percentage valve

$$C_{v0} = F_{max}\sqrt{\frac{\rho}{\Delta p}} \qquad \text{and so} \qquad \frac{C_v}{C_{v0}} = e^{k(V-100)} \tag{5.81}$$

Setting k to zero will give linear behaviour. By choosing the correct value (in this case 0.05) the best possible match can be achieved to the output conditioning required (see Figure 5.35). The characterisation follows Equation (5.76) with a small adjustment to ensure the flow is zero when the valve is fully shut ($V = 0$). Figure 5.37 illustrates the effect of changing k.

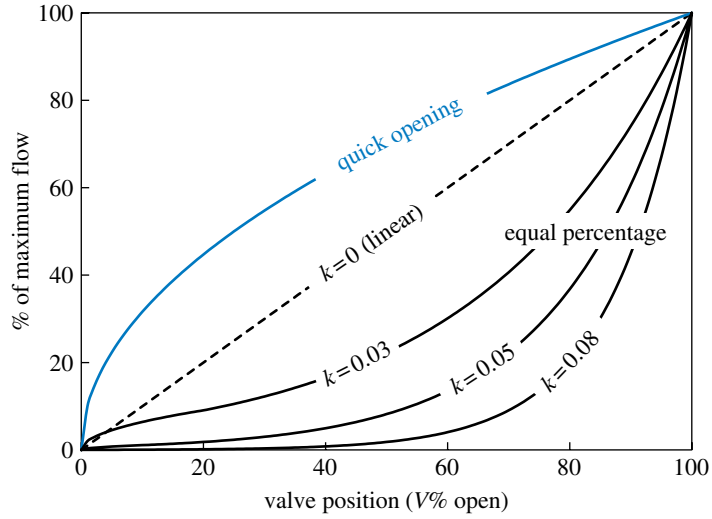

Figure 5.37 *Effect of valve constant*

Equal percentage valves are frequently also used in the control of heat exchangers. If fluid A is used to heat fluid B and there is no change in phase, the energy balance based on flow (F), specific heat (c_p) and temperature (T) can be described simply:

$$F_b\left(c_p\right)_b\left[\left(T_b\right)_{\text{out}}-\left(T_b\right)_{\text{in}}\right]=F_a\left(c_p\right)_a\left[\left(T_a\right)_{\text{in}}-\left(T_a\right)_{\text{out}}\right] \tag{5.82}$$

If we wish to control $(T_b)_{\text{out}}$ by manipulating F_a, the process gain (K_p) can be derived by differentiation:

$$K_p=\frac{d\left(T_b\right)_{\text{out}}}{dF_a}=\frac{\left(c_p\right)_a}{F_b\left(c_p\right)_b}\left[\left(T_a\right)_{\text{in}}-\left(T_a\right)_{\text{out}}\right] \tag{5.83}$$

As the flow of the heating fluid is increased, its outlet temperature approaches its inlet temperature and so the process gain reduces. This nonlinearity can be reduced by having the temperature controller manipulate an equal percentage valve or by applying some similar output conditioning. However, a more precise approach would be to cascade the temperature controller to a flow controller and apply adaptive tuning. The problem can also be resolved more elegantly by the use of a duty controller, as shown in Figure 5.38. With this scheme the MV effectively becomes the exchanger duty and so the process gain is independent of temperature.

$$K_p=\frac{\left(c_p\right)_a}{F_b\left(c_p\right)_b} \tag{5.84}$$

As we will see in Chapter 6, it is also possible to eliminate the dependence of K_p on F_b by changing the MV to a duty-to-flow ratio. However, in the case of exchangers using water in a cooling (or condensing) service, metering of the water flow (F_a) is rarely installed and the use of equal percentage valves is common.

Equal percentage valves are also often installed where they are likely to be over-sized – either because of poor design or because operating flexibility is required. Figure 5.39 illustrates why this is potentially less detrimental than an over-sized linear valve. Both valves have been over-sized by a factor of two. As expected, the maximum opening of the linear valve reduces to 50%. That for the equal percentage valve (with $k = 0.05$) is reduced much less – in this case to about 86%. However, this advantage will be undermined if nonlinearity is introduced unnecessarily. Should an equal percentage valve be installed where a

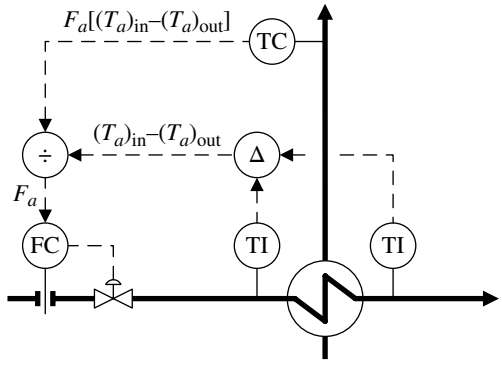

Figure 5.38 *Heat exchanger duty control*

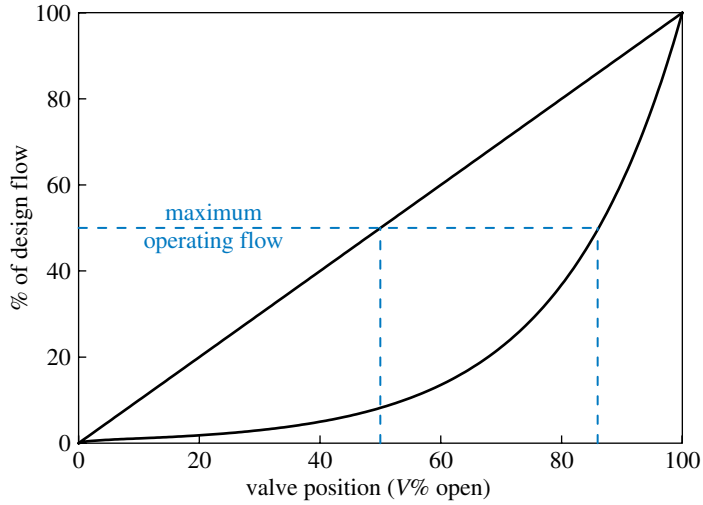

Figure 5.39 *Impact of over-sizing valves*

linear valve would have been better, the valve characterisation can be undone by applying, to the controller output (M), the inverse transformation

$$V = \frac{1}{k}\ln\left(\frac{M}{100}\right) + 100 = 100\left[\frac{\ln\left(M/100\right)}{\ln\left(R\right)} + 1\right] \tag{5.85}$$

It is also possible to apply this same transformation to give a valve a *quick opening* characteristic. Such valves would be selected where the speed of opening is more important than linearity. A common example is their use in anti-surge recycles on compressors.

$$\frac{F}{F_{max}} = \frac{1}{k}\ln\left(\frac{V}{100}\right) + 100 \qquad \text{provided} \quad V > \frac{100}{e^{100k}} \tag{5.86}$$

Again this fails when V is close to zero and so, if the transformation gives a negative result, this is set to zero. Alternatively the square root of the valve position may be used – shown as the coloured line in Figure 5.37.

$$\frac{F}{F_{max}} = 10V^{0.5} \tag{5.87}$$

Quick opening behaviour can also be achieved by transforming the controller output (M) before sending it to a linear valve.

$$V = \frac{100M}{100\left(1-k'\right)+k'M} \qquad \text{where} \quad 0 \le k' < 1 \tag{5.88}$$

Similarly, the inverse function can be applied to compensate for the nonlinear impact of some control valves. For example, the black line in Figure 5.40 shows behaviour typical of a *butterfly* valve. This curve might be available from the manufacturer but more usually is obtained by plotting historically collected flows against the corresponding valve positions. Such valves deliver close to the maximum flow well

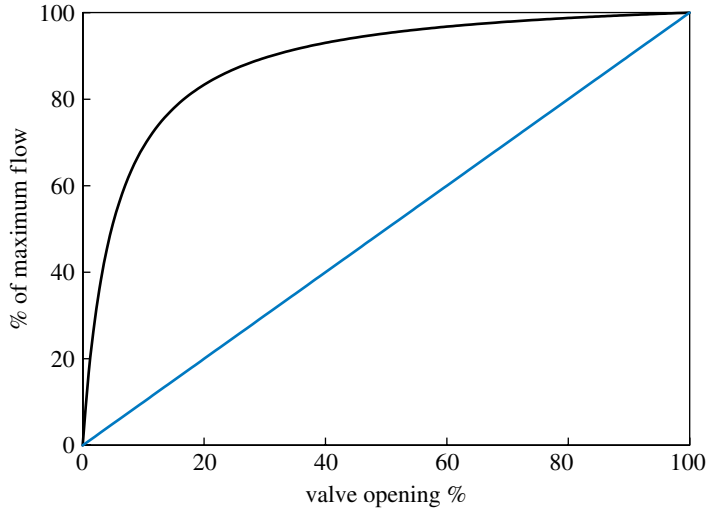

Figure 5.40 *Linearising a butterfly valve*

before being fully opened. While generally not preferred for applications such as flow (and other) controllers, they may be the only economically viable option if the pipe diameter is large. Applying Equation (5.89) with, k' chosen to be 0.95, gives the required linearity as shown by the coloured line.

$$V = \frac{100\left(1-k'\right)M}{100-k'M} \tag{5.89}$$

Note that when developing such output conditioning we have, from the process, historical values of PV and V. We do not have values for M. These must be derived, in this example, by applying the inverse of Equation (5.89), i.e.

$$M = \frac{100V}{100\left(1-k'\right)+k'V} \tag{5.90}$$

In effect, this tells us what M would have been had the conditioning been in place. We then plot the historical values of PV against M – choosing k' to make the relationship as linear as possible.

To emulate equal percentage behaviour, the value for k' can be determined using the short-cut method of forcing the valve characterisation to pass through a chosen point.

$$k' = \frac{100\left(M-V\right)}{\left(100-V\right)M} \tag{5.91}$$

Alternatively several points on the characterisation can be chosen, for example at 10% intervals, and k' chosen to minimise the sum of the squares of deviation from the required curve. The dashed line in Figure 5.35 shows the result of taking this approach – applying Equation (5.89), with k' set to 0.88, to a linear valve. On this occasion this gives a match closer to that required than that achieved with the equal percentage valve. However, valves will rarely behave exactly as characterised and so the best solution must be chosen for each case. Further, providing the chosen technique results in changes of less than ±20% in process gain over the normal operating range of the valve, little noticeable improvement in control performance would be gained by the implementation of a more exact solution.

5.15 Split-Range Valves

While valves are generally calibrated to move over their full range as controller output varies from 0 to 100%, other options are possible. For example, we may wish the valve to fail open on loss of signal – in which case we calibrate it to operate over the range 100 to 0%. We can also calibrate the valve to move over its full range as the controller output changes over only part of its range, e.g. 0 to 50%. We could then calibrate a second valve to move over its full range as the controller output varies from 50 to 100%. If both valves are then manipulated by the same controller, this *split-range* approach would then cause the valves to open and close in sequence. When first applied, split-ranging was implemented by re-calibrating the valve positioners. These days it is more common for the signal to the valve to be conditioned within the control system to achieve the same effect. This simplifies valve maintenance by permitting all valve positioners to be set up in the same way.

Before describing possible applications it is important to distinguish between split-ranging and *dual-acting* valves. Rather than act in sequence, dual acting valves move simultaneously. Figure 5.41 show two ways in which pressure might be controlled in a distillation column. The first scheme controls pressure by changing the condenser duty through manipulation of the bypass. On increasing pressure the controller simultaneously begins to close the bypass and open the valve on the condenser, as shown in Figure 5.42. The valves have been paired to act like a three-way valve – but are less costly and easier to maintain. The second scheme in Figure 5.41 would first open the condenser valve and, if the pressure remains high, begin opening the valve venting vapour from the process. This has been achieved by calibrating the first valve to operate over 0 to 50% of controller output and the second to operate over 50 to 100%, as shown in Figure 5.43.

While split-ranging is common in industry, it does have some limitations. Figure 5.44 shows a method of controlling pressure in a distillation column, often used when vapour production is intermittent. In the absence of sufficient vapour the scheme is designed to allow a non-condensable gas into the column. So, on increasing pressure, the controller will first begin to close valve A until it is fully closed. If the pressure does not fall sufficiently, it will then begin to open valve B. Figure 5.45 shows how the valves have been calibrated. As is common, the range has been split equally between the two valves.

One of the problems with split-ranging is that there can be a large change in process gain as control switches from one valve to the other. To avoid the controller becoming oscillatory it has to be tuned for the range where the process gain is higher. It will thus respond sluggishly when operating in the other part of the range. It is possible to alleviate this problem by redefining the split. The method for doing so is shown

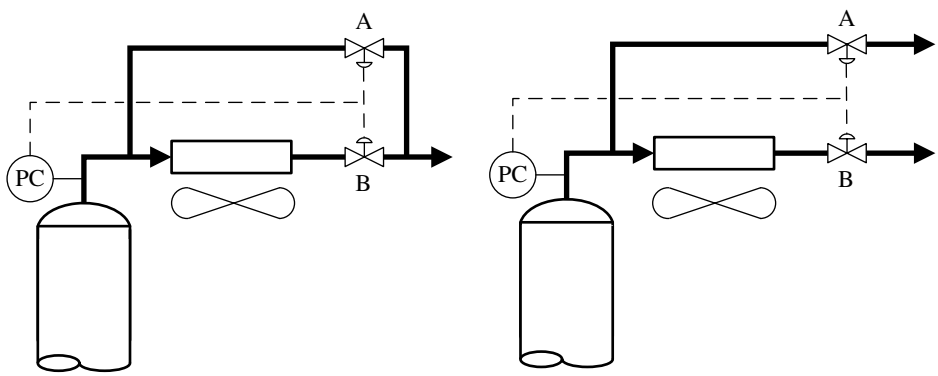

Figure 5.41 *Dual-acting versus split-range*

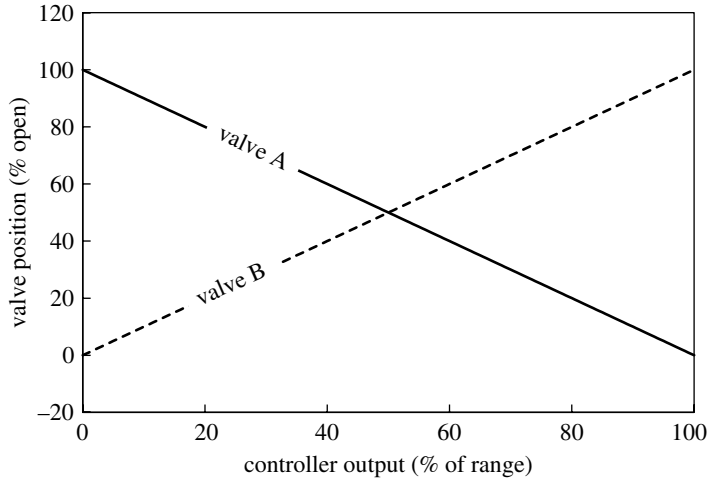

Figure 5.42 *Dual-acting valves*

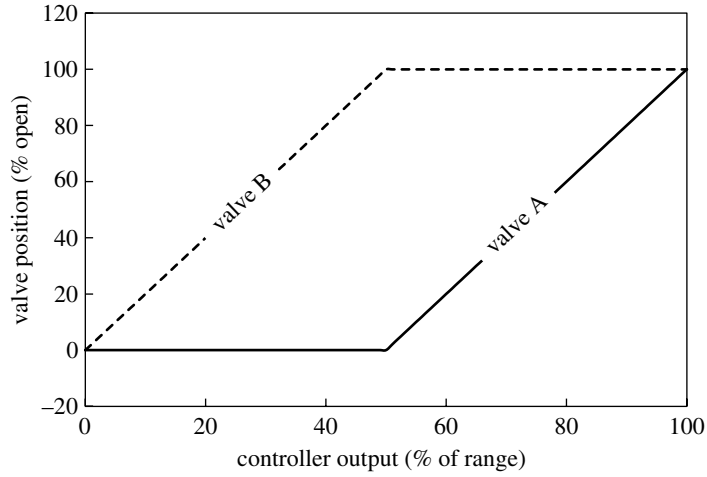

Figure 5.43 *Split-range valves*

in Figure 5.46. The black line represents the process behaviour. It is obtained from historical data or plant tests as described in Chapter 2. In this case it is likely that the pressure will be an integrating process and so, in this example, rate of change of pressure is plotted against controller output. The coloured line is the required relationship. It would be provided by moving the split to around 30%. The same result can be obtained by calculation:

$$OP_{split} = \frac{PV_{split} - PV_{min\ OP}}{PV_{max\ OP} - PV_{min\ OP}}\left(OP_{max} - OP_{min}\right) + OP_{min} \tag{5.92}$$

Referring to Figure 5.46, the minimum OP is 0%; at this point the PV is 95. The maximum OP is 100%; at this point the PV is 10. The PV at the split (when one valve is fully open and the other fully shut) is 70.

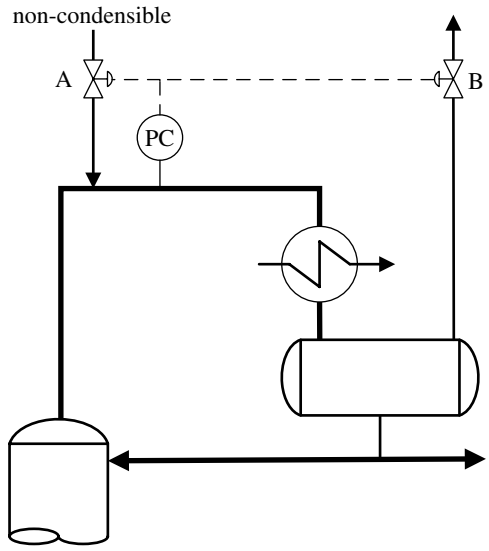

Figure 5.44 *Split-range pressure controller*

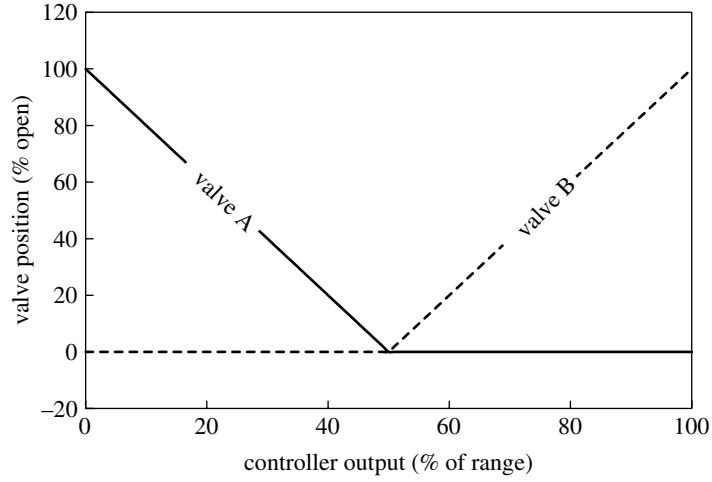

Figure 5.45 *Split-range valve calibration*

Using these values in Equation (5.92) gives a result of 29.4%. Note that in reality it is unlikely that the OP can be varied over the full 0–100% range. The minimum and maximum values used in the calculation should be based on what is achievable on the real process.

While this technique deals with differences in process gain (K_p), it does not account for any difference there might be in the θ/τ ratio. Even with the split optimised, it is possible that the pressure controller will require different tuning as its output moves from the 0–30% part of the range to the 30–100% part. In extreme cases it might not be possible to choose a compromise set of tuning that gives acceptable performance over the whole range.

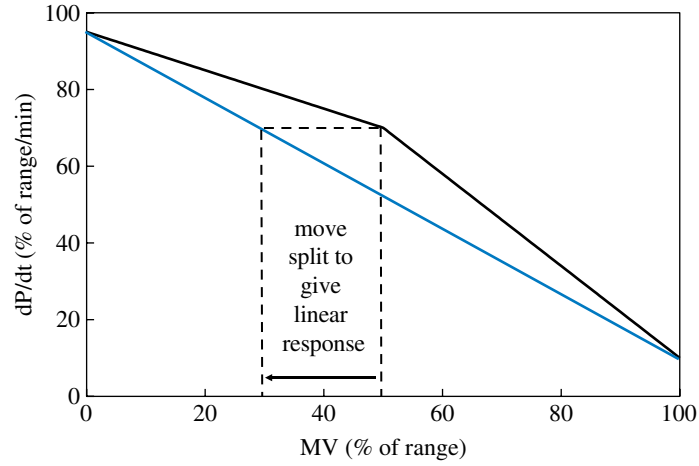

Figure 5.46 *Redefining the split to give a linear response*

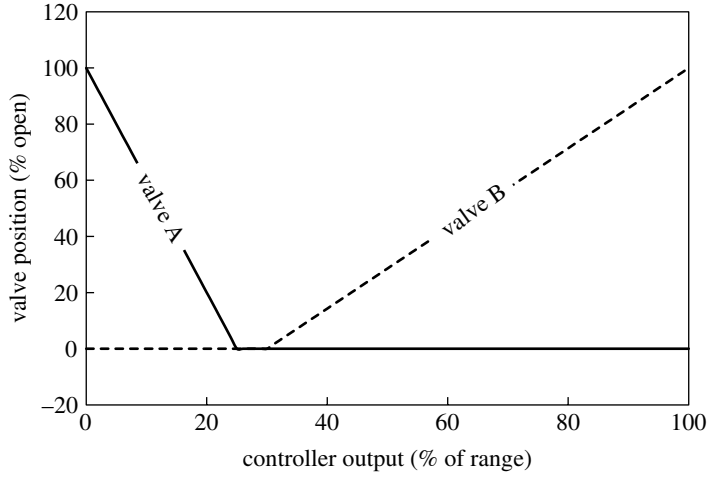

Figure 5.47 *Inaccurate calibration causing a deadband*

A further potential source of problems is the accuracy of the valve calibration. For example if valve A actually travelled its full range as the controller output moved from 0 to 25%, because of some inaccuracy, then there would be a deadband between 25% and 30% where the process gain will be zero – as shown in Figure 5.47. Conversely, if valve A were inaccurately calibrated to operate over the controller output range of 0 to 35%, there would be an *overlap* between 30 and 35% where the process gain is doubled – as shown in Figure 5.48. Both poor calibration situations will severely impact controller performance. Further, if the gas, being used to pressurise the column through valve A, is valuable and depressurisation through valve B is to some waste gas system, then the overlap would mean that both valves are open simultaneously – incurring an unnecessary cost.

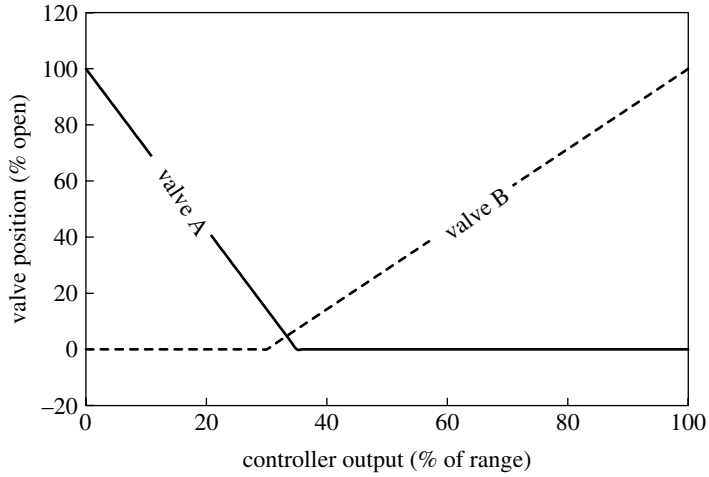

Figure 5.48 *Inaccurate calibration causing an overlap*

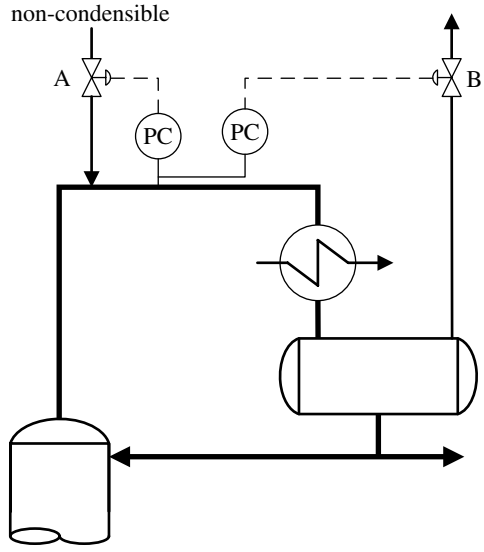

Figure 5.49 *Alternative to split-range pressure controller*

Operators generally interpret a controller output of 0% as a signal to close the valve and 100% to open a valve. With split-ranging, it may not be possible to display the signal in this way and so special attention to operator training will be required. This can be complicated by some control valves being configured as air fail close (or air to open) and others as air fail open (or air to close) depending on what behaviour is required in the event of loss of signal.

In many cases split-ranging is a perfectly satisfactory way of meeting the control objectives. However, poor performance may go unnoticed if the controller spends the majority of the time in one part of the range. It is also possible to split the controller output into more than two parts if there are more MVs

available to extend the operating range. With more splits, accurate valve calibration becomes more important and it is more likely that there will greater variation in process dynamics.

With pneumatic control systems one advantage of split-ranging is that both valves can be connected to the same pneumatic signal line, thus saving the cost of the second line. With multicore cabling or networked 'smart' systems, the incremental cost of the second connection is small. A cost-effective alternative approach is then to design separate controllers, as shown in Figure 5.49. In our example we would have one pressure controller manipulating valve A and another manipulating valve B. Both valves are calibrated normally. This allows us to tune the controllers independently, allowing for any difference in dynamics. It is important that the controllers share the same PV; using independent transmitters with even the slightest measurement error will cause the controllers to fight each other. To achieve the required sequential operation, the controller manipulating valve B would have a slightly higher SP.

6

Feedforward Control

Figure 6.1 shows a simple feedback scheme. The objective is to control the temperature (T) of the mixture of two streams of temperatures T_a and T_b by manipulating the flow of stream B. We have no control over the flow of stream A – indeed, changes in its flow are the main disturbance to the process. The feedback scheme is limited in that it can only take corrective action once it has detected a deviation from temperature SP. This is particularly important if there is significant deadtime between the PV and the MV, for example, if the temperature measurement was a long way downstream of the mixer. No matter how well tuned the feedback controller, it cannot have any impact on the PV until the deadtime has elapsed. During this time the error (E) will increase to

$$E = \left(K_p\right)_a \Delta F_a \left[1 - e^{-\theta_b / \tau_a}\right] \tag{6.1}$$

$(K_p)_a$ is the process gain and τ_a the lag of the temperature in response to a change in the flow of stream A (ΔF_a). θ_b is the deadtime of the temperature in response to its MV, the flow of stream B.

In *feedforward* control we introduce an additional measurement known as the disturbance variable (DV). In this case we can incorporate the measurement of the flow of stream A. By monitoring this

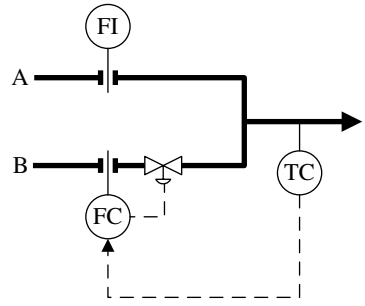

Figure 6.1 *Feedback control*

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

flow we can predict that it will disturb the temperature and so can take action before the temperature changes.

If there was a flow controller on stream A (rather than just an indicator) with its SP being set by, for example, an upstream level controller then this gives us the option of instead using the flow controller's SP as the DV. It could have the advantage of being noise-free and would also permit tighter control by giving an earlier indication of a disturbance. This approach also means that the feedforward controller will not take unnecessary action if a load change is being handled by the DV controller. For example, a change in upstream pressure will change the flow of stream A but will be dealt with by the flow controller. Any move made by a feedforward controller, based on flow PV, would need to be reversed later. Logic might be incorporated into the scheme to handle the situation where the DV controller does not properly respond to SP changes if, for example, the controller is saturated. But, if the DV controller is switched to manual mode, PV tracking will mean that the DV will automatically switch to using the measurement of the controller – although this may introduce noise and slightly change the feedforward dynamics.

Under some circumstances it may be necessary to disable feedforward control, for example, because of an instrument failure. This is usually done by switching to a constant for the DV, freezing it at the last good value. This graceful degradation reverts bumplessly to the feedback scheme without intervention from the operator or any need to adjust tuning constants. Care needs to be taken in recommissioning feedforward. The DV is likely to have moved away from the frozen value and the feedback controller will have taken corrective action. Simply re-enabling feedforward will then bump the process. To avoid this the controller would usually be forced to reinitialise. Alternatively enabling feedforward should only be permitted if the feedback controller is in manual mode.

6.1 Ratio Algorithm

Figure 6.2 shows a possible feedforward scheme employing a *ratio algorithm*. This algorithm generates an output by multiplying the two inputs. One input is the measured flow of stream A; the other is the operator-entered target for the ratio of the flow of stream B to stream A. As the flow of stream A changes, the

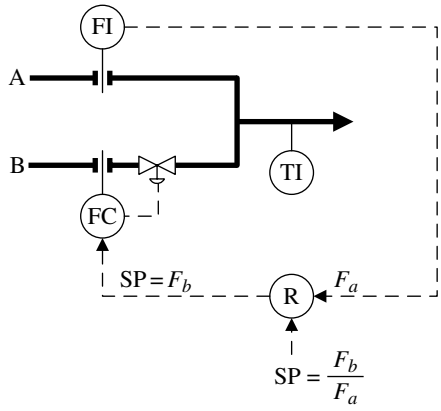

Figure 6.2 *Feedforward control*

ratio algorithm maintains the flow of stream B in proportion, thus keeping the temperature constant. If the liquids have the specific heats of $(c_p)_a$ and $(c_p)_b$ then, in order to meet the target temperature (T), this ratio (R) will be set as:

$$R = \frac{(c_P)_a}{(c_p)_b}\left[\frac{T - T_a}{T_b - T}\right] \tag{6.2}$$

The ratio algorithm is available in most DCS as a standard feature. Strictly, since it does not have feedback, it is not a controller – although it is often described as such. It is, however, more than just a simple multiplier. It incorporates the equivalent of PV tracking. When in manual mode the ratio SP tracks the actual ratio. If R is the ratio SP, I the input measurement and M the algorithm's output, the algorithm performs the calculation

$$R = \frac{M}{I} \tag{6.3}$$

When switched to automatic the ratio SP is fixed at the current value and the calculation changes to

$$M = R \times I \tag{6.4}$$

Once initialised in this way the operator may change the ratio SP, or may cascade a controller to adjust it as necessary. This means that we do not need to measure the stream temperatures and specific heats, as used in Equation (6.2). Provided that, when the ratio is switched to automatic, the target temperature is being met then the ratio SP will be automatically initialised to the correct value. Further, to make it appear like a controller, the DCS will permit both the ratio SP and the actual ratio to be displayed. In some DCS, the ratio algorithm is incorporated into the PID controller and can be configured as an option.

A less preferred approach, occasionally described in other texts, is shown in Figure 6.3. The ratio is calculated from the two flow measurements and used as the PV of a true ratio controller. If we assume that F_b is linearly related (with constant of proportionality k) to the valve position, set by the controller output (M), then

$$R = \frac{F_b}{F_a} = \frac{kM}{F_a} \qquad \text{and} \qquad K_p = \frac{dR}{dM} = \frac{k}{F_a} \tag{6.5}$$

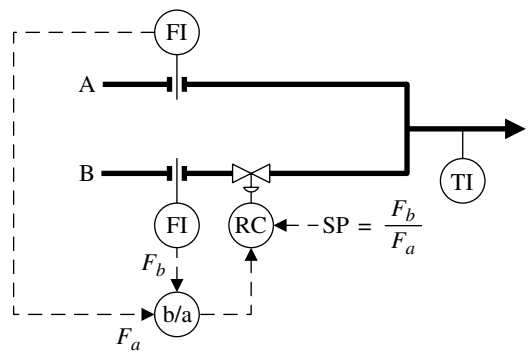

Figure 6.3 *Alternative configuration for feedforward control*

The process gain of the ratio with respect to changes in valve position is now inversely proportional to the uncontrolled flow. In theory the ratio controller gain should be kept in proportion to the uncontrolled flow – but only if the variation is significantly greater than ±20%.

Feedforward is not a replacement for the feedback scheme. It does not incorporate all possible disturbances. For example, it would not compensate for a change in the temperature of either stream. While in theory we could include these measurements as additional DVs, it is unlikely that we could include all possible sources of disturbance. For instance, measuring changes in the specific heat of either liquid is unlikely to be practical. Further, feedforward relies on the instrumentation being accurate. If there was a bias error in the measurement of either flow, maintaining a fixed proportion would not meet the target temperature. Finally, we need to provide the process operator with a practical way of changing the target temperature. It should not be necessary to manually recalculate the revised target ratio.

We therefore treat feedforward as an enhancement to feedback control. We combine them to give a *feedforward-feedback* scheme where, in this case, the temperature controller now manipulates the ratio target. However, in this particular case, which is an example of *blending*, a better approach is to manipulate the ratio between the controlled flow and the total flow. If F_a and F_b are the flows of the two streams then, if the specific heats of the two streams are the same

$$F_a T_a + F_b T_b = \left(F_a + F_b\right) T \tag{6.6}$$

If we define the ratio (R), as shown in Figure 6.2, as F_b/F_a:

$$T = \frac{T_a + RT_b}{R+1} \qquad \text{and} \qquad K_p = \frac{dT}{dR} = \frac{T_b - T_a}{\left(R+1\right)^2} \tag{6.7}$$

This shows that K_p is not constant. For example, if R is typically 0.5 but varies outside the range 0.35 to 0.65, K_p will vary by more than ±20% and cause problems with tuning the temperature controller. However, if we define R as $F_b/(F_b + F_a)$:

$$T = \left(1 - R\right) T_a + R.T_b \qquad \text{and} \qquad K_p = \frac{dT}{dR} = T_b - T_a \tag{6.8}$$

The process gain no longer varies with R. So, if the ratio is likely to be highly variable, the feedforward scheme shown as Figure 6.4 would be preferable. Figure 6.5 shows the full feedforward-feedback scheme. However, process gain will still vary with temperature. If T_a and T_b are numerically similar then a relatively small change in either will cause a proportionately large change in the difference between them and so cause a large change in K_p. Provided the temperatures are measured, then some form of adaptive tuning could be readily implemented. However, a better approach would be to replace the volumetric flows, used in the ratio algorithm and PID controller, with enthalpy flows calculated from the temperatures. This would not only ensure that K_p remains constant but also that temperature changes will be dealt through the feedforward scheme rather than the feedback temperature controller.

A common application of ratio control is in blending two or more liquid components to meet a product quality specification. If blending into a product tank, the algorithm used for this application will usually be based on totalised flows so that if the required ratio cannot be achieved, for example because a flow controller saturates, the shortfall can be compensated for later. Alternatively the scheme may include *pacing* which reduces the total product flow until any saturation is alleviated.

In the same way that ratio algorithms can be used to combine streams in the required proportion, they can also be used to divide a stream. For example it may be required to distribute a feed stream between several parallel units of different capacity.

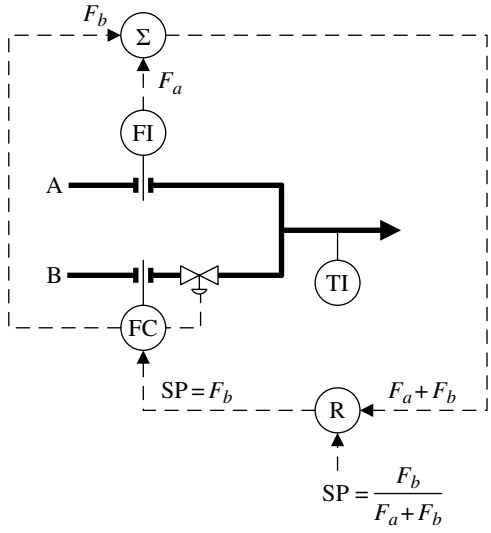

Figure 6.4 *Preferred blend ratio control*

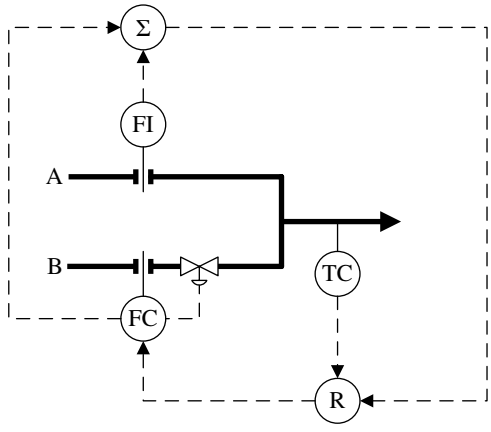

Figure 6.5 *Feedforward-feedback control*

6.2 Bias Algorithm

The ratio algorithm provides a means by which two strategies can manipulate the same variable. In our example both the feedforward and feedback parts of the scheme can change the flow of stream B as necessary without fighting each other. Rather than multiply them together, an alternative means of combining two signals is to add them. So, in addition to the ratio algorithm, the DCS is likely to include a *bias algorithm*. This performs in a way similar to the ratio. When in manual mode the bias (B) is determined by

$$B = M - I \tag{6.9}$$

When switched to automatic this calculation is replaced by

$$M = I + B \tag{6.10}$$

Using a bias algorithm would be incorrect in our mixing example. The temperature would not be kept constant by fixing the total (or the difference) of the two flows. However, there are occasions where the use of the bias is correct and the ratio not so. Figure 6.6 shows one such situation. The heater burns a combination of two fuels. The first (fuel A), perhaps a by-product from another part of the process, can vary. The second (fuel B) can be manipulated to control the heater outlet temperature. Without feedforward in place, any change in the flow of fuel A would disturb the temperature. Given that the process is likely to have a significant deadtime the disturbance will not be detected immediately by the temperature controller. And, for the same reason, the controller will not support very fast tuning. The deviation from SP is therefore likely to be large and lengthy.

With the flow of fuel A incorporated as a DV, we can apply the bias algorithm to maintain the total fuel constant. The algorithm supports the addition of a scaling factor on its input. In this case we set this factor to −1. If we think of the output from the temperature controller being the total fuel required, the bias now subtracts the flow of fuel A from this and the result is the SP of the fuel B flow controller. Thus any change in the flow of fuel A causes an immediate compensation in fuel B and the temperature will remain constant. In Chapter 10 we show how this technique can be applied even if the fuels are quite different, for example, if one is a gas and the other a liquid.

As with ratio feedforward, some DCS include the bias as an option within the PID algorithm. Some also permit configuration of the operator display to make the algorithm appear like a true controller – showing both the actual bias and the bias SP. An alternative approach is shown in Figure 6.7. The two fuel flow measurements are summed, with suitable scaling factors to ensure the result is in units consistent with the total energy supplied. The result is then used as the PV of a total energy flow controller.

In this dual-fired heater example it would be incorrect to apply a ratio algorithm. We do not wish to keep the two fuel flows in proportion. We have already shown that we can use only ratio feedforward for blending. However, there are situations where either algorithm may be used. For example, in Chapter 4, we described how the three-element level control scheme for a steam drum may be adapted to either approach. Similarly, in the same way that a ratio algorithm can be used to distribute a stream, so can the bias algorithm.

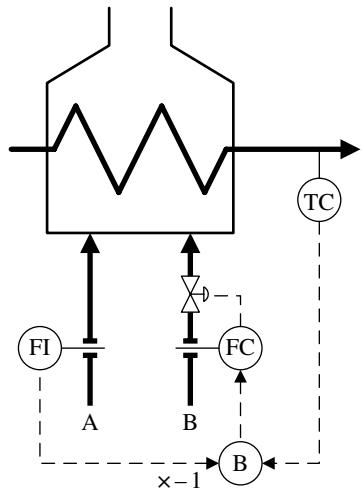

Figure 6.6 *Bias feedforward example*

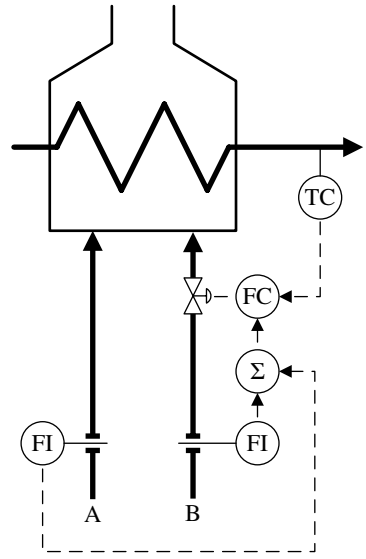

Figure 6.7 Alternative configuration for bias feedforward

6.3 Deadtime and Lead-Lag Algorithms

So far we have ignored process dynamics; we have assumed that the feedforward controller should change the MV at the same time as the DV changes. This is only correct if the dynamics of the PV with respect to changes in the MV are the same as its dynamics with respect to the DV. In our examples this is the case, but this is not always so. Consider the modification made to the blending process shown in Figure 6.8. A surge drum has been added, fitted with an averaging level control, so that fluctuations in the flow of stream A are reduced. However, the temperature control scheme is unchanged.

An increase in the flow of stream A will cause an immediate increase in the flow of stream B. But, because the additional stream A is largely accumulated in the drum and not routed to the mixer, this will cause the combined temperature to change. The feedback controller will detect this and will bring the temperature back to SP by reducing the ratio target and hence the flow of stream B. However, the drum level controller will ultimately increase the flow of stream A to the mixer, causing a second disturbance to the temperature. The temperature controller will compensate for this by bringing the ratio SP back to its starting point.

The feedforward controller therefore made the correct change; it just did so too soon. The dynamics of the PV with respect to the DV are now much slower than its dynamics with respect to the MV. We therefore need to include some dynamic compensation that, in this case, delays the feedforward correction. Failure to properly include such compensation can result in the addition of feedforward causing the scheme to perform less well than the standalone feedback controller.

Of course, this is a contrived example. Had the process design department consulted the plant's control engineer, the drum would have been installed upstream of the measurement of the flow of stream A. Further, the averaging level controller could then properly be cascaded to a flow controller.

The dynamic compensation required uses algorithms that are provided by the DCS. These are the *deadtime algorithm* and the *lead-lag algorithm*. The deadtime algorithm generates a pure delay, just like process deadtime, except the delay is a tuning constant (θ) configurable by the engineer. The lead-lag

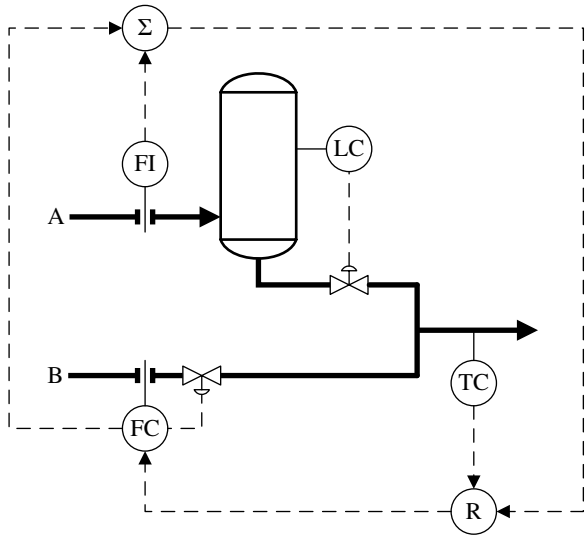

Figure 6.8 *Need for dynamic compensation*

algorithm has three tuning constants: gain (K), lead ($T1$) and lag ($T2$). If X is the input and Y the output then, in the time domain, it has the form

$$Y = K\left(1 + \frac{T1 - T2}{T2} e^{-t/T2}\right) X \tag{6.11}$$

As we show in Chapter 15, the discrete combined deadtime/lead-lag algorithm should be of the form

$$Y_n = e^{-ts/T2} Y_{n-1} + K\frac{T1}{T2} X_{n-\theta/ts} - K\left(\frac{T1 - T2}{T2} + e^{-ts/T2}\right) X_{n-\theta/ts-1} \tag{6.12}$$

Generally θ will not be an exact multiple of ts and so the values of $X_{n-\theta/ts}$ and $X_{n-\theta/ts-1}$ are linearly interpolated between adjacent values of X.

$$X_{n-\theta/ts} = X_{n-\text{int}(\theta/ts)} - \left(\frac{\theta}{ts} - \text{int}\left(\frac{\theta}{ts}\right)\right)\left(X_{n-\text{int}(\theta/ts)} - X_{n-\text{int}(\theta/ts)-1}\right) \tag{6.13}$$

$$X_{n-\theta/ts-1} = X_{n-\text{int}(\theta/ts)-1} - \left(\frac{\theta}{ts} - \text{int}\left(\frac{\theta}{ts}\right)\right)\left(X_{n-\text{int}(\theta/ts)-1} - X_{n-\text{int}(\theta/ts)-2}\right) \tag{6.14}$$

The algorithm is coded in different ways in different DCS. For example, to avoid the use of the exponential function, we can assume that $ts << T2$ and so make the first order Taylor approximation

$$e^{-ts/T2} = 1 - \frac{ts}{T2} \tag{6.15}$$

And so we obtain

$$Y_n = \frac{T2 - ts}{T2} Y_{n-1} + K\frac{T1}{T2} X_{n-\theta/ts} - K\left(\frac{T1 - ts}{T2}\right) X_{n-\theta/ts-1} \tag{6.16}$$

Or we can apply the reciprocal Taylor approximation:

$$e^{-ts/T2} = \frac{1}{e^{ts/T2}} = \frac{1}{1+\frac{ts}{T2}} \tag{6.17}$$

And so obtain:

$$Y_n = \frac{T2}{T2+ts}Y_{n-1} + K\frac{T1}{T2}X_{n-\theta/ts} - K\left(\frac{T1}{T2} - \frac{ts}{T2+ts}\right)X_{n-\theta/ts-1} \tag{6.18}$$

We can also apply the more accurate first order Padé approximation:

$$e^{-ts/T2} = \frac{2-\frac{ts}{T2}}{2+\frac{ts}{T2}} \tag{6.19}$$

And so we get:

$$Y_n = \frac{2.T2-ts}{2.T2+ts}Y_{n-1} + K\frac{T1}{T2}X_{n-\theta/ts} - K\left(\frac{2.T1}{2.T2+ts}\right)X_{n-\theta/ts-1} \tag{6.20}$$

We showed in Chapter 3, for the PID algorithm, there are several versions – depending on the method used to convert the analog form into its closest discrete equivalent. The same applies to the lead-lag algorithm. As with the PID algorithm, if the scan interval (ts) is small compared to the tuning constants, the differences in output will be indistinguishable. Few DCS vendors disclose exactly how the algorithm is coded. The approach to tuning must therefore be to assume that it is theoretically correct and, if it does not behave as exactly as expected, to adjust the tuning constants by trial-and-error.

Figure 6.9 shows the effect of the algorithm. The dashed line shows a step change in input (ΔDV). The gain term (K) determines the steady state change in output. The impact of adjusting K is shown in Figure 6.10. The output is delayed by θ. Figure 6.11 shows the impact of adjusting θ. The output then changes as a step with the height of the step determined by the ratio of $T1$ to $T2$. Figure 6.12 shows

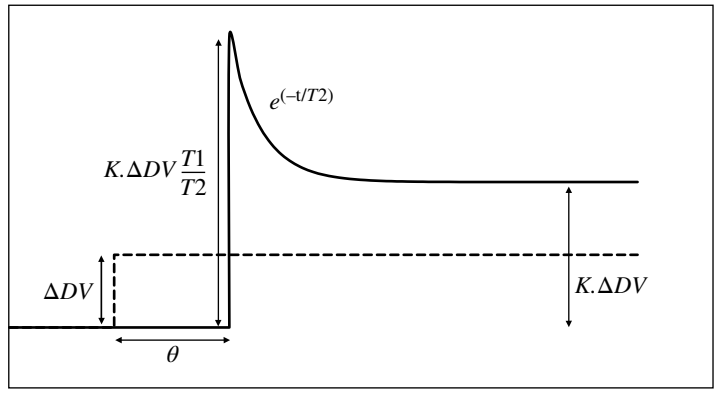

Figure 6.9 *Effect of deadtime and lead-lag algorithms*

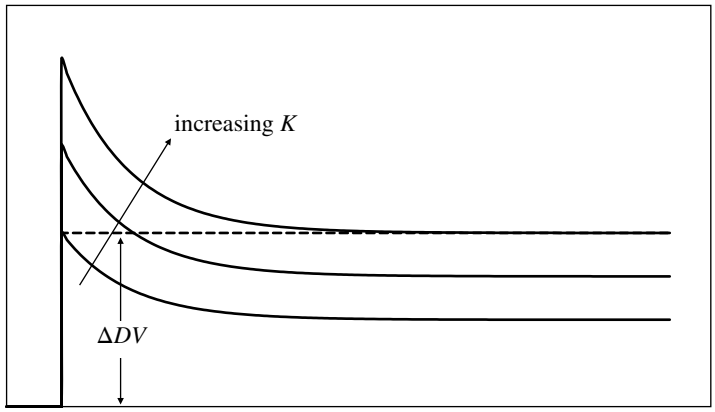

Figure 6.10 *Effect of K*

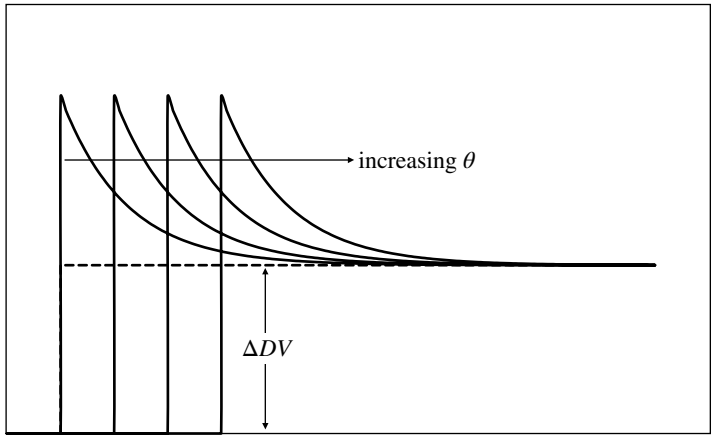

Figure 6.11 *Effect of θ*

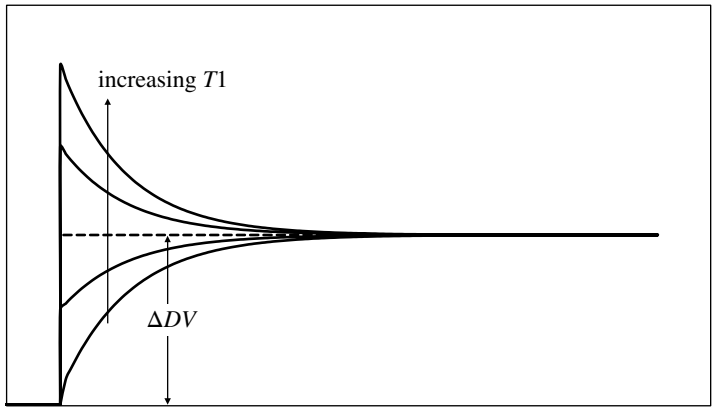

Figure 6.12 *Effect of T1*

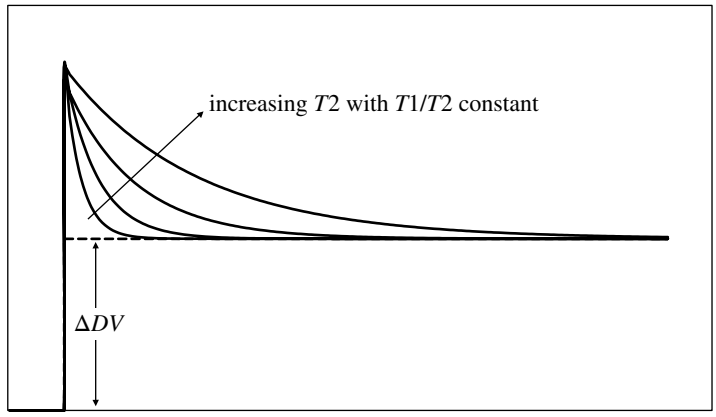

Figure 6.13 *Effect of T2*

the effect of varying the lead term ($T1$). With $T1$ set to zero the step is eliminated. If $T1$ is less than $T2$ then the step is less than the steady state change, if greater than $T2$ then the output overshoots the steady state change. After the step the output approaches the steady state condition with a lag of $T2$. Figure 6.13 shows the effect of adjusting $T2$ – simultaneously adjusting $T1$ to maintain the ratio.

If $T1$ is equal to $T2$ the lead and lag exactly cancel and the step in the input passes through the algorithm as a step. By also setting K to 1 and θ to zero the output of the algorithm will be the same as the input.

6.4 Tuning

The inclusion of this algorithm has added four tuning constants. When feedforward is added to an existing feedback scheme, we will show later that it may be necessary also to re-tune the PID controller. Tuning potentially seven constants by trial-and-error would be extremely time-consuming. While a little fine tuning may be necessary, we should use the process dynamics to obtain the best possible estimate. Ideally we should be able to obtain an estimate which works first time.

As an example we will add a feedforward ratio scheme to our case study heater. Its schematic is shown in Figure 6.14. The principle behind the scheme is that it will maintain a constant fuel-to-feed ratio. This is not a blending problem, as described in Figure 6.3, and so this definition of the ratio will not give tuning problems. Indeed, we shall see later that it resolves one. From a steady state point of view, holding the ratio constant is a good approximation. It is unlikely to be perfect because the heater efficiency will change a little as feed rate is changed. However, the feedback controller will correct for this by trimming the ratio SP.

Without dynamic compensation, the scheme will immediately change the fuel rate as the feed rate changes. If the temperature responds more quickly to fuel changes than it does to feed changes, then the correction will have been made too soon and the temperature will show a transient deviation from SP. The feedback controller does not ‘know’ that the temperature will eventually return to its starting point and will take corrective action. This unnecessary action will later result in another temperature disturbance.

We need first to check whether dynamic compensation is necessary and, if so, obtain estimates for the tuning constants. The approach, as usual, is to first develop a full understanding of the process dynamics.

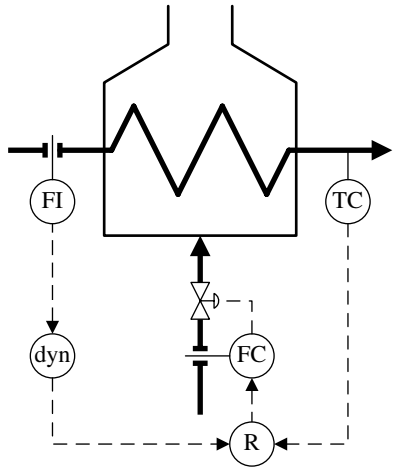

Figure 6.14 *Feedforward-feedback schematic*

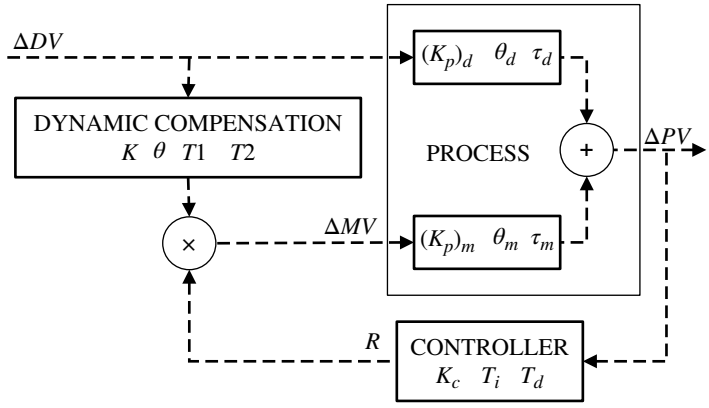

Figure 6.15 *Feedforward-feedback block diagram*

By step-testing the fuel flow SP (MV) we obtain the dynamic behaviour of the temperature (PV). We may already have these dynamics from steps conducted to tune the PID controller. These dynamics we define as $(K_p)_m$, θ_m and τ_m. In addition we need the dynamics for the DV. With the temperature controller in manual, we step the feed rate (DV) and obtain the dynamic behaviour of the temperature – giving us $(K_p)_d$, θ_d and τ_d.

In general we can draw this scheme as the block diagram shown in Figure 6.15. The objective of the scheme is to ensure there is no change in heater outlet temperature ($\Delta PV = 0$) when the feed rate is changed. For this to be the case, the temperature change caused by ΔDV must be exactly balanced by the change caused by ΔMV, i.e.

$$\Delta DV.\left(K_p\right)_d = \Delta DV.K.R.\left(K_p\right)_m \tag{6.21}$$

By definition

$$\left(K_p\right)_m = \frac{\Delta PV}{\Delta MV} \qquad \left(K_p\right)_d = \frac{\Delta PV}{\Delta DV} \qquad R = \frac{\Delta MV}{\Delta DV} \tag{6.22}$$

Substituting these into Equation (6.21) shows that K should be set to 1. This is always the case for ratio feedforward. For bias feedforward:

$$K = -\frac{\left(K_p\right)_d}{\left(K_p\right)_m} \tag{6.23}$$

Figure 6.16 shows the effect of this design in response to an increase in feed rate. This would normally cause the temperature to decrease as shown but the increase in fuel rate (made at the same time by the ratio controller) causes a compensating increase in temperature – the net effect of which should be to keep the temperature constant. While this is effective at steady state, there is a transient change in temperature caused by the mismatch in process dynamics. In this example we can see that the deadtime for the change in fuel is about three minutes less than that for the change in feed rate; so increasing fuel at the same time as feed causes the temperature to initially increase. The transient is also a function of the difference in the lags of the two changes. The feedback controller will respond to this disturbance but its action is unnecessary – worsening the disturbance. Therefore, in addition to making a feedforward adjustment of the correct magnitude, we must also make it at the correct time. Firstly, it must have the same deadtime, and so

$$\theta_d = \theta + \theta_m \qquad \text{or} \qquad \theta = \theta_d - \theta_m \tag{6.24}$$

Secondly, it must have the same lag. The lag between temperature and fuel is τ_m while that between temperature and feed is τ_d. We first cancel out τ_m this by setting the lead term ($T1$) equal to this value and then replace it with τ_d by setting the lag term ($T2$) equal to this value.

In summary, for ratio feedforward, we have tuning as follows:

$$K = 1 \qquad \theta = \theta_d - \theta_m \qquad T1 = \tau_m \qquad T2 = \tau_d \tag{6.25}$$

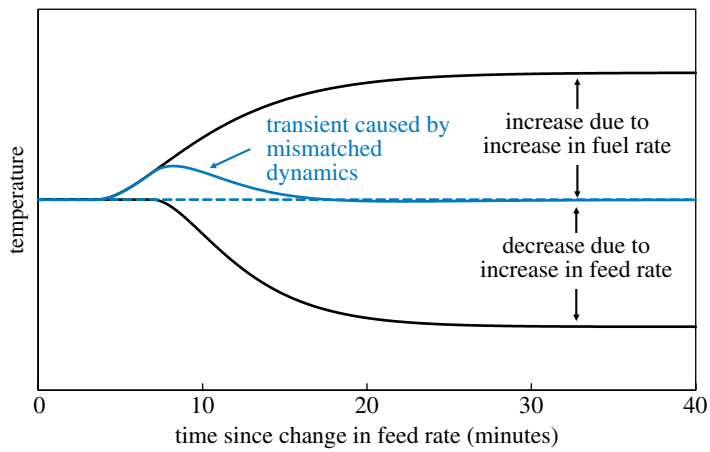

Figure 6.16 *Mismatch of process dynamics*

The effect of these parameters would make, in Figure 6.16, the temperature trend caused by the change in fuel rate the mirror image of that caused by the change in feed rate.

An alternative derivation of the tuning method, using Laplace transforms, is presented at the end of this chapter.

Before implementation, there are a number of checks to make. Firstly, there is no guarantee that θ_d is greater than θ_m. Thus θ can be negative; this means that we would have to change the fuel before the feed rate changes. Such a requirement is described as not *realisable*. As a compromise, θ is set to zero and $(\theta_m - \theta_d)$ is added to $T1$. We effectively make up for delaying the fuel change by increasing the 'spike'.

Of course, if θ is close to zero and $T1$ is close to $T2$, the dynamic compensation may be omitted from the configuration because the dynamic response of the PV to the DV is the same as that to the MV. However, this may not always be the case. It may be worthwhile including the algorithm, but setting θ to zero and $T1$ equal to $T2$. This would permit compensation to be added easily if required.

We need to consider any noise that may be present in the DV. In our case this was previously only an indication and so any noise was not passed through to the control valve. This will no longer be the case and so filtering may be necessary. Ideally the filter should be put in place before step-testing but, if this has been overlooked, we can compensate for its addition by increasing $T1$ by the filter time constant (τ_f).

No matter how fast the response to the DV, $T2$ should not be set to zero. This would cause a full-scale kick to the MV when the DV changes. Indeed, we should check the $T1/T2$ ratio in any case. For example, if this is greater than 1.15, the MV overshoot will be greater than 15%. We may need to reduce $T1$ and make the same compromise that we do for PID control, i.e. accept a slower return to SP in order to avoid excessive changes to the MV. If θ is nonzero, it is possible to partially compensate for the reduction in $T1$ by reducing θ.

We should also remember that the tuning has been based on the assumption that the process is first order plus deadtime. It is theoretically possible to implement a second order equivalent of the lead-lag algorithm but this would require the identification of second order models for the DV and MV, and the calculation of additional tuning constants. It is unlikely therefore to be practical. It would be easier to fine tune the dynamic compensation. This also takes account of any abnormalities in the way in which the DCS vendor may have coded the lead-lag algorithm.

Tuning needs to be approached systematically before the feedback controller is commissioned. Otherwise we could be simultaneously adjusting up to seven tuning constants. If bias feedforward has been configured then the value of K can be adjusted by examining the steady state behaviour. If, after a change to the DV, the PV returns to its starting value, then the value for K is correct. If not, then by determining whether feedforward has undercompensated or overcompensated, K should be increased or decreased. However, if the process is nonlinear, adjusting K can improve the performance of the controller for changes in one direction but worsen it in the other. If ratio feedforward has been selected, as we have seen, K should be fixed at 1. Any adjustment will result in the displayed ratio differing from the true ratio.

Once K is correct, θ is next adjusted as required. This is determined by considering how the PV would have changed with no feedforward in place. If, when the DV changes, the PV initially moves in the same direction as it would without feedforward, then θ should be reduced. If the PV moves in the opposite direction then θ should be increased. The next stage is to adjust $T1$, following the same method – but increasing it if the PV moves in the same direction as the open loop response, decreasing it otherwise. $T2$ should be reduced if the PV appears to only slowly return to steady state – remembering that $T1$ should also be changed to maintain a constant $T1/T2$ ratio.

It is common that ratio feedforward is added to an existing feedback controller. If this is the case then we should check whether any change is necessary to its tuning. Since it is now manipulating the ratio SP,

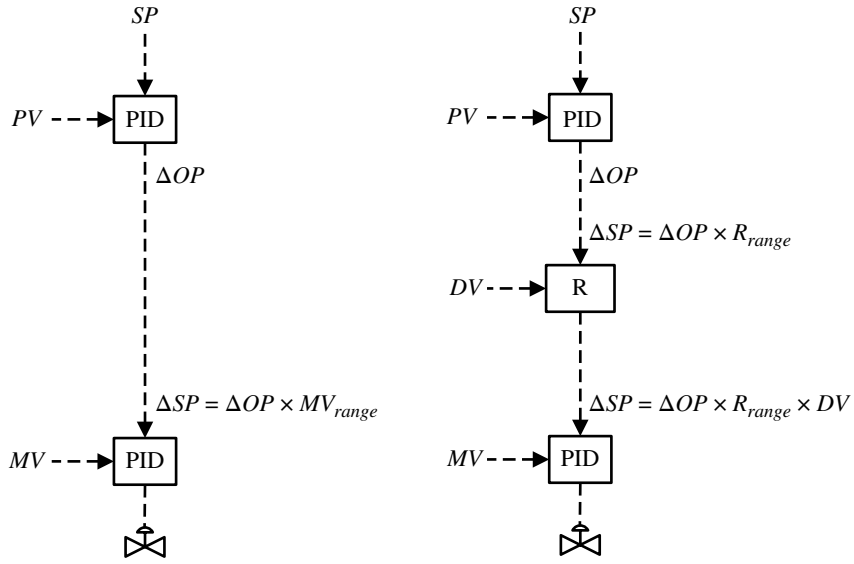

Figure 6.17 *Change of control configuration*

rather than the fuel SP, we may have to compensate for the fact that we have changed the range of its MV. Figure 6.17 shows the effect of the change.

While internally controllers operate over a dimensionless range, ratio and bias algorithms normally work in engineering units. Prior to the implementation of feedforward the PID controller output was multiplied by the range of the MV (MV_{range}) to convert it to engineering units. After implementation it is multiplied by the range of the ratio (R_{range}) and by the DV. This will change the loop gain; to compensate, we therefore have to adjust the controller gain.

$$\left(K_c\right)_{new} = K_c \times \frac{MV_{range}}{R_{range} \times DV*} \tag{6.26}$$

Of course, DV is not a constant; however, it is correct to use $DV*$ – the value of DV at which $(K_p)_m$ was determined.

We have some control over this correction factor since we choose the range of the ratio when configuring it in the DCS. Ideally we would like the factor to be 1 and so

$$R_{range} = \frac{MV_{range}}{DV*} \tag{6.27}$$

This not only means that we do not have to re-tune any existing feedback controller but it also means that we can permit the operator to switch off the feedforward if, for example, there is an instrument problem and retain the feedback scheme without re-tuning. However, we do have to set the range to accommodate the lowest and highest ratios that might be expected during normal operation. This is given by

$$R_{range} = \frac{\text{maximum value of MV}}{\text{minimum value of DV}} - \frac{\text{minimum value of MV}}{\text{maximum value of DV}} \tag{6.28}$$

If the result from Equation (6.28) is larger than that from Equation (6.27) then this must take precedence and K_c recalculated using Equation (6.26).

Of course, an existing feedback controller may benefit from re-tuning in any case. It may be configured with the wrong algorithm, such as proportional-on-error rather than proportional-on-PV, or it may be that attention has never been paid to optimising its tuning. New tuning can be derived from the methods presented in Chapter 3, using $(K_p)_m$, θ_m and τ_m and modifying the resulting K_c according to Equation (6.26). Alternatively new dynamics could be obtained by commissioning only the feedforward scheme and stepping the ratio SP. This would then take account of any change of instrument range.

In addition to more quickly responding to process disturbances, ratio feedforward on feed rate offers another, less immediately obvious, advantage. We showed in Chapter 2 that process gain is usually inversely proportional to feed rate. This means, that as feed rate changes, we should adjust controller gain in proportion to feed rate in order to keep the loop gain constant. Examination of Figures 6.14 and 6.15 shows that, with feedforward in place, the PID controller output is multiplied by feed rate – effectively increasing the controller gain in proportion, as required. The performance of the feedback controller will therefore be the same at any feed rate. In Chapter 2 we showed that controllers on a process with a turn-down ratio greater than 1.5 are likely to need re-tuning as feed rate changes. Under these circumstances ratio feedforward on feed rate should be considered, even if feed rate disturbances are relatively rare, since this would avoid the need to re-tune.

It should also be noted that nowhere in our tuning calculations for ratio feedforward is the value of $(K_p)_d$ used. There is thus no need to take account of any changes that might occur in its value because, for example, the process is highly non-linear. In fact its value can even change sign, as we illustrate in Chapter 12, without causing any control problem. $(K_p)_m$, although not used in calculating the feedforward tuning, does influence feedback tuning. If it changes for reasons other than changes in feed rate then we may have to apply some form of adaptive tuning.

Process gains on integrating processes do not change with feed rate. The use of ratio feedforward would therefore bring a disadvantage in that the feedback controller would require re-tuning if feed rate changes by more than 20%. If the use of either bias or ratio feedforward makes good process sense, as it did for the steam drum level controller described in Chapter 4, then bias feedforward would have the advantage.

6.5 Laplace Derivation of Dynamic Compensation

One of the key aims of this book is to present the subject using the minimum of mathematics. Laplace transforms are a very convenient way of representing dynamic behaviour but can be daunting to control engineers unfamiliar with this branch of mathematics. We therefore have largely avoided using them, but include here an example of how they can be used effectively.

The disturbance made to the PV as a result of the change in DV is given by

$$PV_d = \frac{\left(K_p\right)_d e^{-\theta_d s}}{\tau_d s + 1} DV \tag{6.29}$$

The transform for the dynamic compensation comprises a gain (K), deadtime (θ) and lead-lag ($T1$ as lead, $T2$ as lag). It has the form

$$MV = Ke^{-\theta s} \frac{T1s + 1}{T2s + 1} DV \tag{6.30}$$

The change in PV caused by the feedback action is

$$PV_m = \frac{\left(K_p\right)_m e^{-\theta_m s}}{\tau_m s + 1} MV \tag{6.31}$$

For there to be no change in PV:

$$PV_d + PV_m = 0 \tag{6.32}$$

Or

$$\frac{\left(K_p\right)_d e^{-\theta_d s}}{\tau_d s + 1} = -\frac{\left(K_p\right)_m e^{-\theta_m s}}{\tau_m s + 1} K e^{-\theta s} \frac{T1s + 1}{T2s + 1} \tag{6.33}$$

Equating coefficients gives the tuning for bias feedforward.

$$K = -\frac{\left(K_p\right)_d}{\left(K_p\right)_m} \qquad \theta = \theta_d - \theta_m \qquad T1 = \tau_m \qquad T2 = \tau_d \tag{6.34}$$

By definition:

$$\left(K_p\right)_d = \frac{\Delta PV}{\Delta DV} \qquad \left(K_p\right)_m = \frac{\Delta PV}{\Delta MV} \tag{6.35}$$

Therefore:

$$K = -\frac{\Delta MV}{\Delta DV} \tag{6.36}$$

For ratio feedforward, $(K_p)_d$ and $(K_p)_m$ must have opposite signs and so K is positive. By definition the SP (R) of the algorithm is the ratio of the MV to the DV. So the algorithm already provides the necessary feedforward gain term and K should be set to 1.

7

Deadtime Compensation

We have seen in Chapter 3 that as the θ/τ ratio increases we have to substantially reduce the controller gain (K_c) to maintain stability. Thus, not only is there a delay in first detecting the disturbance, the controller can only respond slowly. We therefore are likely to see large and sustained deviations from SP. A number of techniques have been published which help overcome this problem. However, they rely on estimates of the process dynamics and only offer an advantage if these can be quantified with reasonable accuracy. If poor estimates are used or if the dynamics change unpredictably then performance can degrade to less than what can be achieved with conventional PID control.

7.1 Smith Predictor

Perhaps the earliest, and most well-known technique, is that developed by Smith [23]. It still employs a PID controller but in addition includes a *fast model* that predicts how the process will behave in the future. The controller uses the output of this model, rather than the actual PV, and can therefore be tuned as if there is no deadtime. A *plant model* is also included. Its purpose is to check whether the actual PV eventually matches the prediction and, if not, generate a correction term. Figure 7.1 shows the configuration.

The fast model is simply a standard lag block. Its gain is set to the process gain and its lag set to the process lag. Its input is the output of the PID controller. Thus, when the controller takes corrective action, the output of the fast model will begin changing immediately. As far as the controller is concerned the θ/τ ratio is zero. It can then therefore be tuned very tightly – in theory only restricted by any MV overshoot limit. Of course, it is not sufficient merely to control a model of the process. We have to ensure that the fast model truly represents future process behaviour. This is the purpose of the plant model. It is the same as the fast model but, in addition, includes deadtime set equal to the process deadtime. Rather than predicting the future it now models the current process behaviour. Because it is otherwise the same as the fast model, it can be used to compare what the fast model predicted against what is actually happening now. Figure 7.2 shows an alternative configuration where the plant model uses the fast model with the addition of a delay set to the process deadtime. The difference between the actual PV and the predicted ('model') PV is added to the output of the fast model as a correction term. The PID controller should be tuned as usual but remembering the θ/τ ratio is now zero so little derivative action will be required.

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

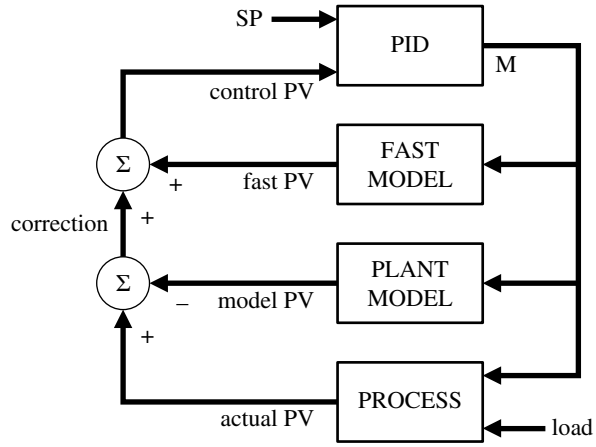

Figure 7.1 *Configuration of the Smith predictor*

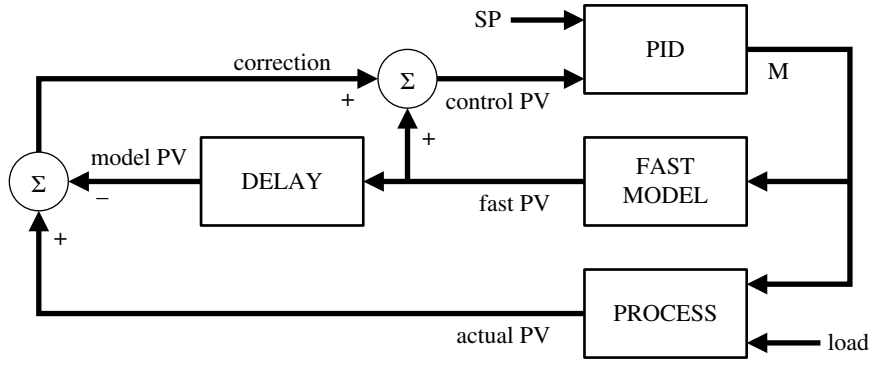

Figure 7.2 *Alternative configuration of the Smith predictor*

The Smith is an example of a *predictor-corrector* controller. The output of each part of the model, in response to an open loop change in the PID controller output, is shown in Figure 7.3. The trend for the control PV shows a typical transient disturbance immediately the process deadtime has elapsed. This comes from the corrector and is caused by a mismatch between the actual and the predicted PV. It might arise because of a difference between the process deadtime and the deadtime in the plant model. It might also be caused by the plant model being first order plus deadtime whereas the process probably has a higher order. It is this transient that can limit how large a gain may be used in the PID controller.

Figure 7.4 shows the closed loop behaviour. Compared to an optimally tuned conventional PID controller, the Smith predictor performs considerably better. In fact, it can be tuned much more aggressively – limited only by the accuracy of the model and the MV overshoot. While the MV shows no overshoot, its trend however, does show regular disturbances caused by modelling error. The controller responds to these as if they are load disturbances and the action it takes later, when the deadtime has elapsed, cause a disturbance in the actual PV. This will then repeat at an interval equal to the process deadtime. If the controller gain is too high each transient will be greater than the one before and the controller will become unstable.

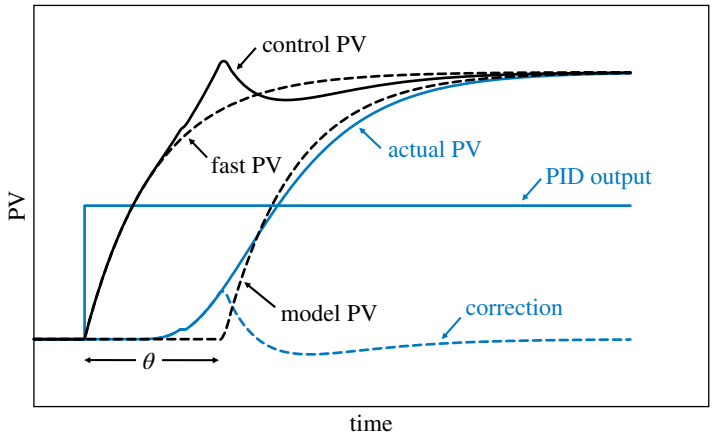

Figure 7.3 *Outputs in Smith predictor*

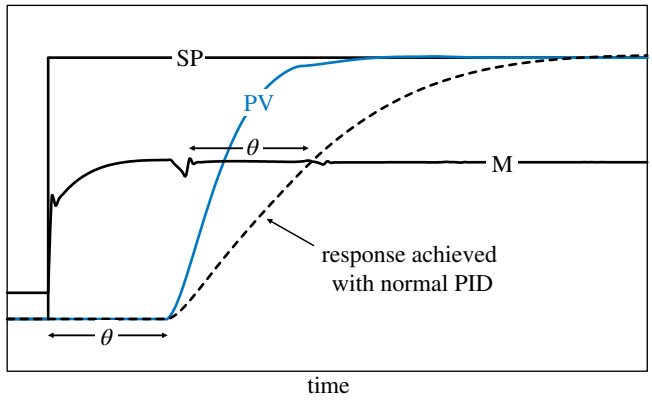

Figure 7.4 *Performance of Smith predictor*

We described in Chapter 2 how the correct choice of deadtime, used in the preferred tuning methods described in Chapter 3, results in a very robust PID controller. Figure 7.5 shows the impact of the process deadtime falling by about 20% without a correction made to controller tuning – typically what might occur if there was an increase in feed rate. The Smith predictor now performs far worse than the PID controller, emphasising the importance of not applying the technique unless the process dynamics are accurately known. However, if the dynamics change predictably, adaptive tuning can be applied to the plant model – for example by making its deadtime a function of feed rate. If only deadtime varies then no changes need to be made to the tuning of the fast model or the PID controller.

There are a number of ways in which the Smith predictor may be modified. In addition to deadtime, the time constant can be removed from the fast model. It will then become a simple gain term and therefore also compensates for lag. However, removing it completely will make indeterminate the θ/τ ratio as 'seen' by the controller. Most tuning methods require this ratio. Setting τ instead to a small value will assist with tuning the PID.

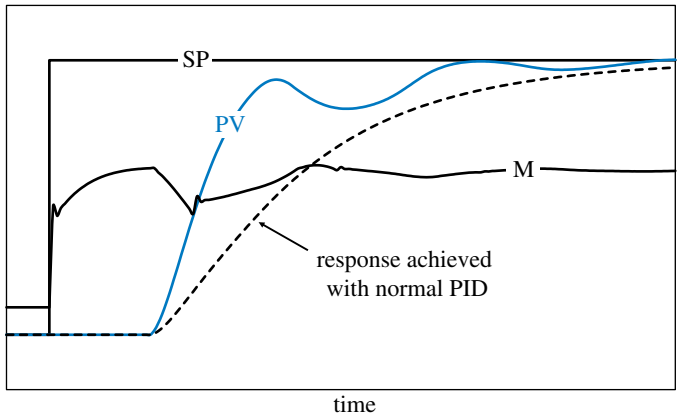

Figure 7.5 Performance of Smith predictor with deadtime error

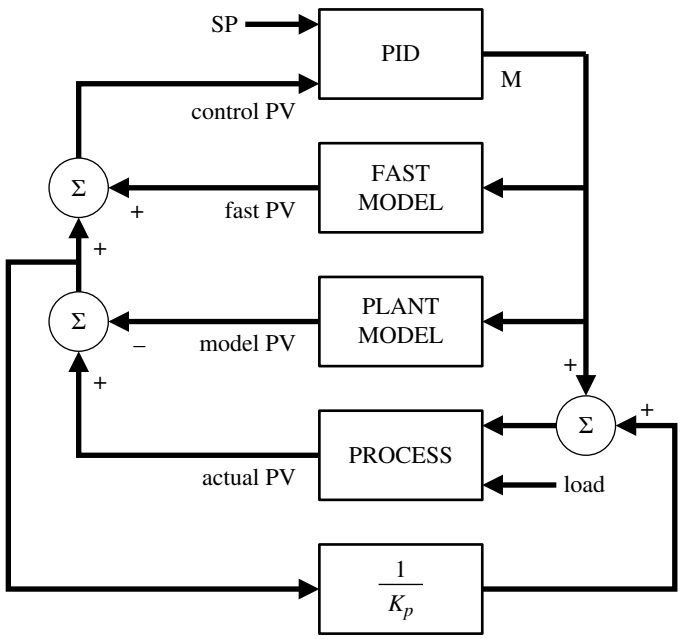

Figure 7.6 Modified Smith predictor

Another possible enhancement is shown in Figure 7.6. The correction term is now multiplied by the reciprocal of the process gain before being added to the controller output. In principle this has the effect of immediately compensating for the model error rather than this being performed by the PID controller. However, since the PID controller is now tuned to act more quickly, the benefit, as shown in Figure 7.7, is relatively small. But since load changes are also a cause of model error, directionally these too will be dealt with more quickly. We describe in the next section how this approach can be used to replace, rather than simply supplement, the PID controller.

It is also possible to change to second order or higher models, provided the additional time constant(s) can be accurately determined. For example the Smith predictor can in principle be applied to processes exhibiting inverse response. As described in Chapter 2, inverse response arises when there are two competing sub-processes. Assuming zero deadtime, as a simplification, a process might have a process gain of $(K_p)_1$ and a lag of τ_1 with a competing process, causing the inverse response, having a (negative) gain of $(K_p)_2$ and lag of τ_2. For the inverse response to be apparent, τ_2 would be significantly less than τ_1. Referring to Figure 7.8 the fast model is a lag set to τ_2 with a gain of $(K_p)_2$. Its output is added to the real PV, theoretically cancelling the inverse response. The dynamics of the control PV should therefore appear to be first order with a gain of $(K_p)_1$ and lag of τ_1. The PID controller is tuned assuming these dynamics. The process model is a second order representation of the process and is used to generate the model correction. Techniques for identifying second order models with inverse response are described in Chapters 14 and 15. However, it can be difficult to achieve accuracy sufficient for the controller to be robust.

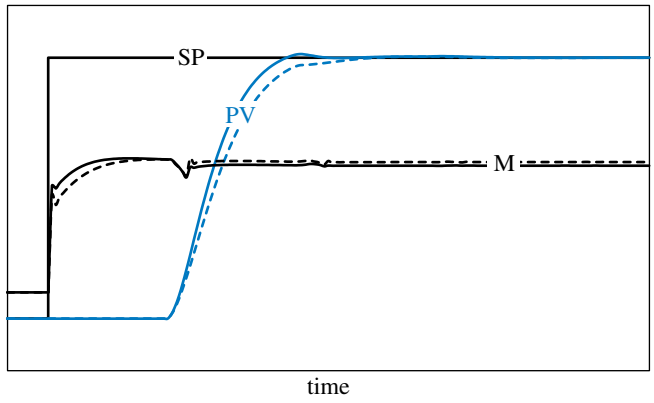

Figure 7.7 *Effect of modification to Smith predictor*

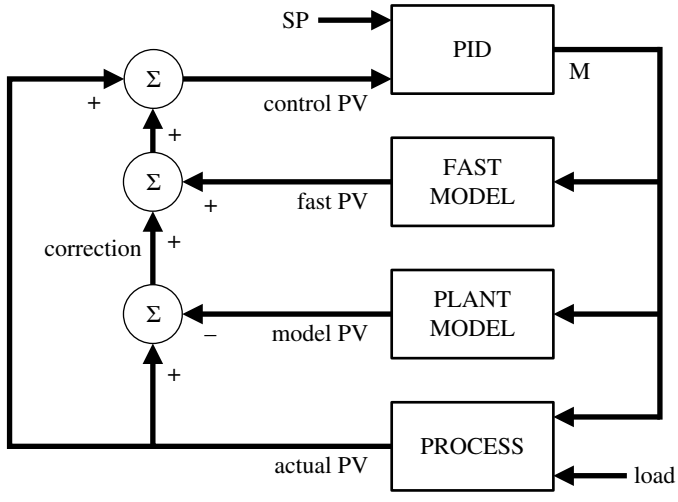

Figure 7.8 *Smith predictor configures to compensate for inverse response*

7.2 Internal Model Control

IMC (internal model control) replaces the PID controller completely. We saw in Chapter 3 that there is also an IMC tuning method for PID control. Tuned using this method the performance of the PID is theoretically the same and its replacement. However, this is only exactly true if there is no deadtime. With deadtime IMC will outperform IMC tuning in a PID controller. The scheme has also been described as *dynamic reconciliation* [21].

Figure 7.9 shows its configuration. Any changes in SP are divided by the process gain (K_p) and pass via a lead-lag algorithm to the process. The lead term ($T1$) is set to the process lag. The lag term ($T2$) is configured to give the desired controller response. If set equal to the process lag, the change is sent to the process as a step. The PV will then approach the SP with a time constant equal to the open loop process lag (τ). If $T2$ is set to a value less than τ, a more aggressive change is made and the MV will overshoot its steady state value. Increasing $T2$ above τ will result in a slower approach to SP with no MV overshoot.

If the gain is accurate then no further control action is required. If not then there will be a mismatch between the process and the plant model. This model is used in the same way as the Smith controller, to generate a correction term which, in this scheme, is subtracted from the SP.

Performance is equivalent to that of the Smith predictor and it too will perform badly if there is significant model error, particularly with deadtime. As with the Smith predictor, higher order models can be used if required.

7.3 Dahlin Algorithm

The Dahlin algorithm [24] follows the form

$$MV_n = a_0E_n + a_1E_{n-1} + \ldots + b_1MV_{n-1} + b_2MV_{n-2} + \ldots \tag{7.1}$$

This is a generalised version of the control algorithm. For example, by setting

$$a_0 = K_c\left[1+\frac{ts}{T_i}+\frac{T_d}{ts}\right] \qquad a_1 = -K_c\left[1+2\frac{T_d}{ts}\right] \qquad a_2 = K_c\frac{T_d}{ts} \qquad b_1 = -1 \tag{7.2}$$

We get

$$\Delta MV = K_c\left[\left(E_n - E_{n-1}\right)+\frac{ts}{T_i}E_n+\frac{T_d}{ts}\left(E_n - 2E_{n-1}+E_{n-2}\right)\right] \tag{7.3}$$

This is the equation for the proportional-on-error, derivative-on-error ideal controller. We can, however, choose coefficients to produce almost any control algorithm. For example the *deadbeat controller* is defined as

$$MV_n = a_0E_n + a_1E_{n-1} + b_1MV_{n-(N+1)} \qquad \text{where} \qquad N = \frac{\theta}{ts} \tag{7.4}$$

The coefficients a_0, a_1 and b_1 are chosen to ensure that, following a SP change, the error becomes zero when one scan interval after the deadtime has elapsed and remains zero for all subsequent scans. This is equivalent to minimising IAE but it is possible to determine the coefficients without following the normal

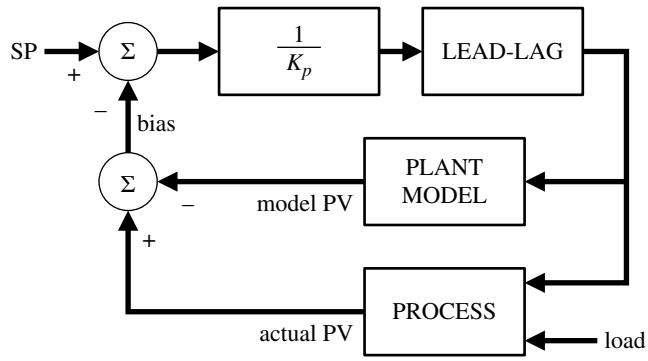

Figure 7.9 *Configuration of IMC*

iterative search procedure. By examining Equation (7.4) we can see that b_1 must be 1 because at steady state all values of E are zero and all values of MV are the same.

From Equation (3.126), we can describe a first order process as

$$PV_n - PV_{n-1} = P\left(PV_{n-1} - PV_{n-2}\right) + \left(1-P\right)K_p\left(MV_{n-1} - MV_{n-2}\right) \quad \text{where } P = e^{-ts/\tau} \tag{7.5}$$

From Equation (7.4), if θ is zero

$$MV_{n-1} - MV_{n-2} = a_0E_{n-1} + a_1E_{n-2} = a_0\left(PV_{n-1} - SP_{n-1}\right) + a_1\left(PV_{n-2} - SP_{n-2}\right) \tag{7.6}$$

Substituting in Equation (7.5)

$$PV_n - PV_{n-1} = P\left(PV_{n-1} - PV_{n-2}\right) + \left(1-P\right)K_p\left(a_0\left(PV_{n-1} - SP_{n-1}\right) + a_1\left(PV_{n-2} - SP_{n-2}\right)\right) \tag{7.7}$$

The process is at steady state when n is –1 and, when n is 0, we change the SP from SP_{old} to SP_{new}. So, when n is 1

$$PV_1 - PV_0 = P\left(PV_0 - PV_{-1}\right) + \left(1-P\right)K_p\left(a_0\left(PV_0 - SP_{new}\right) + a_1\left(PV_{-1} - SP_{old}\right)\right) \tag{7.8}$$

The PV will remain unchanged until n is 1 and so PV_{-1} and PV_0 will be equal to SP_{old}. We require E_1 to be zero and so PV_1 will be equal to SP_{new}. Therefore,

$$a_0 = \frac{-1}{K_p\left(1-P\right)} = \frac{-1}{K_p\left(1-e^{-ts/\tau}\right)} \tag{7.9}$$

When n is 2

$$PV_2 - PV_1 = P\left(PV_1 - PV_0\right) + \left(1-P\right)K_p\left(a_0\left(PV_1 - SP_{new}\right) + a_1\left(PV_0 - SP_{new}\right)\right) \tag{7.10}$$

We require E_2 to be zero and so PV_2 will be equal to SP_{new}. Therefore,

$$a_1 = \frac{P}{K_p\left(1-P\right)} = \frac{e^{-ts/\tau}}{K_p\left(1-e^{-ts/\tau}\right)} \tag{7.11}$$

Equating coefficients between Equations (7.4), (7.9), (7.11) and Equation (3.27) shows that we can configure the PI algorithm as a deadbeat controller with

$$K_c = \frac{-e^{-ts/\tau}}{K_p\left(1-e^{-ts/\tau}\right)} \qquad \text{and} \qquad T_i = \frac{e^{-ts/\tau}}{\left(1-e^{-ts/\tau}\right)}ts \tag{7.12}$$

The negative controller gain arises from defining E as $PV–SP$. It means that, if K_p is positive, the controller should be configured as reverse acting.

By modifying Equation (3.126) we can describe a first order process with deadtime as

$$PV_n - PV_{n-1} = P\left(PV_{n-1} - PV_{n-2}\right) + \left(1-P\right)K_p\left(MV_{n-(N+1)} - MV_{n-1-(N+1)}\right) \tag{7.13}$$

If θ is not zero then, in general, a deadbeat controller becomes

$$MV_n = \frac{-1}{K_p\left(1-e^{-ts/\tau}\right)}E_n + \frac{e^{-ts/\tau}}{K_p\left(1-e^{-ts/\tau}\right)}E_{n-1} + MV_{n-(N+1)} \tag{7.14}$$

This cannot be converted to an equivalent PID algorithm.

Equation (3.130), predicting MV overshoot, was developed for the same controller as that described by Equation (7.12). This showed that, for the overshoot to be less than the 15% limit we would typically apply, we can only apply deadbeat control to processes that have dynamics of around 0.5 second. Dahlin addressed this problem by modifying the technique to include the required time constant (λ) for the trajectory of the approach to SP, as used in the Lambda tuning method described in Chapter 3. For a first order plus deadtime processes we use

$$MV_n = a_0E_n + a_1E_{n-1} + b_1MV_{n-1} + b_{N+1}MV_{n-(N+1)} \tag{7.15}$$

where

$$a_0 = \frac{-\left[1-e^{-ts/\lambda}\right]}{K_p\left[1-e^{-ts/\tau}\right]}. \qquad a_1 = \frac{e^{-ts/\tau}\left[1-e^{-ts/\lambda}\right]}{K_p\left[1-e^{-ts/\tau}\right]} \tag{7.16}$$

$$b_1 = e^{-ts/\lambda} \qquad b_{N+1} = 1-e^{-ts/\lambda} \qquad N = \frac{\theta}{ts} \tag{7.17}$$

Note that, because we define error as $PV–SP$, the signs of the coefficients a_0 and a_1 are opposite to those that are generally published.

It is unlikely that θ will be an exact multiple of ts and so the value of $MV_{n-(N+1)}$ is linearly interpolated between adjacent values of MV.

$$MV_{n-(N+1)} = MV_{n-\text{int}(N+1)} + \left(N-\text{int}\left(N\right)\right)\left(MV_{n-\text{int}(N+1)} - MV_{n-\text{int}(N+2)}\right) \tag{7.18}$$

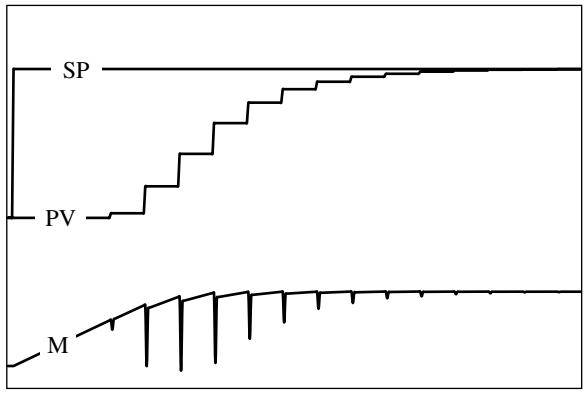

Figure 7.10 *Derivative spikes caused by discontinuous measurement*

The performance of the Dahlin algorithm is similar to that of the Smith predictor and IMC. It is equally sensitive to the accuracy of the deadtime (θ) used in deriving N and hence the chosen value of $MV_{n-(N+1)}$. It too can be extended to higher-order models.

The algorithm was originally developed for use when the controller scan interval (ts) is significant compared to the process dynamics. This makes it suitable for use if the PV is discontinuous, such as that from some types of on-stream analysers. Analysers are a major source of deadtime. They may be located well downstream of the MV and their sample systems and analytical sequence can introduce a delay. An optimally tuned PID controller would then have a great deal of derivative action. However, this will produce the spiking shown in Figure 7.10. We covered derivative spikes in Chapter 3 and, by switching derivative action to be based on PV rather than error, were able to eliminate them when the SP was changed. However, this change has no effect if the step change in error arises because of a similar change in PV. There are, however, several solutions.

- Simply retuning the controller as PI only would eliminate the spikes but the PV would return to SP more slowly.
- Filtering the measurement would smooth the steps but the level of filtering required would change the overall dynamics, requiring the PID controller to be re-tuned – probably with a much reduced gain.
- No derivative action is required if the Smith predictor is used. However, the discontinuities will cause a mismatch between the process and the plant model, which in turn will cause spurious corrections. IMC will have a similar problem.

The most elegant solution is to take control action only when the analyser generates a new reading – effectively making the controller scan interval the same as the analyser sample interval. Most analysers have a *read-now* contact that can be brought into the DCS and used to initiate a control scan. However, the analyser sample interval may not be constant. While some operate with a timed sequence, others will only move to the next step of the sequence once the previous one is complete. Steps like cooling or heating to a required temperature can vary in duration. We should not then apply a technique like PID which assumes a fixed sample interval. Techniques, such as Dahlin, readily permit the coefficients to be based on the actual sample interval. This also has the advantage that control action is only taken if the analyser produces a new reading. Relying only on the PV from the analyser's sample-and-hold logic will mean that control action will continue even if the measurement is not being refreshed, causing the controller eventually to saturate.

8

Multivariable Control

8.1 Constraint Control

As the name suggests, constraint controllers are designed to drive the process towards operating limits in a direction known to be profitable. Constraints are either hard or soft. Hard constraints are those which can only be approached from one side. This might be because it is mechanically impossible; for example a control valve cannot open beyond 100%. Or a hard constraint might be imposed for safety reasons; for example operating pressure can theoretically be increased above a relief valve setting but doing so would be extremely hazardous. Finally, there may be strong operational reasons for not violating a constraint; for example minimising the flow of recycle around a compressor should not cause the machine to surge. While the machine itself might tolerate this for some period, the loss of gas flow to the process would cause severe operational problems.

Soft constraints can be approached from either side. Violation of a soft constraint does not cause a major problem, provided corrective action is taken promptly. An example is the quality of liquid products leaving a process. If the product is ultimately routed to storage then small amounts of off-grade production can be mixed with material with *giveaway*. Another example is liquid level; while low and/or high limits may be set, these may be violated briefly with no impact on the process. A maximum limit is applied to skin (tube metal) temperatures on some fired heaters. This may be either to keep coking at a reasonable level or to prevent the tubes from *creeping*. Both are long-term cumulative effects and are unlikely to be measurably worse if the maximum temperature is occasionally violated.

Some MPC packages permit the definition of hard and soft constraints. In addition to their meaning as described above, often a soft constraint is used as a more conservative limit on a hard constraint. MPC will violate soft constraints if this is the only way that it can satisfy all the hard constraints.

Constraint controllers fall into three categories.

- Single-input single-output (SISO) controllers, as the name suggests, comprise controllers that can manipulate only one MV to approach only one constraint.
- Multi-input single-output (MISO) controllers are required when there is more than one constraint that is approached by adjusting just one MV. Such controllers require some form of logic to select the most limiting constraint. Once selected the controller behaves just like a SISO controller.
- Multi-input multi-output (MIMO) controllers adjust multiple MVs in order to satisfy multiple constraints. Generally, although not necessarily, the controller can reach as many constraints as there are MVs. Thus,

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

if there are more constraints than MVs, some selection is necessary. MPC packages generally incorporate a *linear program (LP)* for this purpose. This permits some simple economics to be defined by the engineer to drive the controller to select the most profitable set of constraints to approach. This establishes the SPs. More than one constraint will be affected by each MV; otherwise we would only require a set of SISO controllers. Thus the controller must handle the interactions to ensure that approaching one constraint does not cause another to be violated. If there are fewer constraints than MVs then the controller will have *degrees of freedom* – generally equal to the number of MVs less the number of constraints. There will be multiple ways in which the constraints can be approached. A single-input multi-output (SIMO) is a special example of this situation. Selection of the most profitable strategy becomes an economic optimisation problem, usually outside the scope of MPC – although most support optional add-ons that address this.

Care must be taken when determining the number of degrees of freedom. Figure 8.1 shows a 2 × 2 constraint control problem. Given that it is possible to simultaneously reach both constraints by adjusting the MVs, it would be tempting to assume that there are no degrees of freedom. In fact, the most profitable operating point lies on only one of the constraints. At this point there is then one degree of freedom.

8.2 SISO Constraint Control

Very simple constraint control is possible using the conventional PID control algorithm. Figure 8.2 shows a fired heater that has a single constraint – a hydraulic limit on fuel. It is profitable to run the heater at maximum feed rate, subject to this constraint.

The first indication that the process is operating at the maximum feed rate is the fuel valve fully opening. Sometimes described as a *valve position controller (VPC)*, a single-input single-output (SISO) controller, with this valve opening as its PV, manipulates the SP of the feed flow controller until the fuel valve reaches the desired maximum opening. On most processes, valve position transmitters are only installed in special circumstances. What is actually used here is the output of the fuel flow controller. We assume that the valve positioner is properly calibrated and working well.

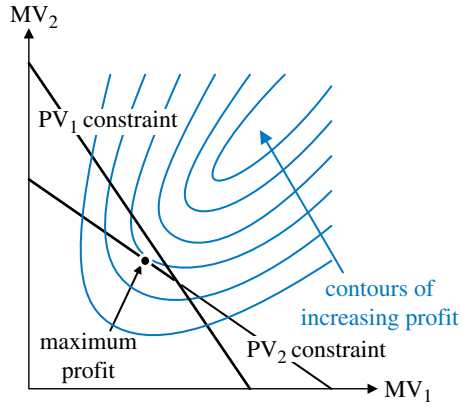

Figure 8.1 *Optimisation problem*

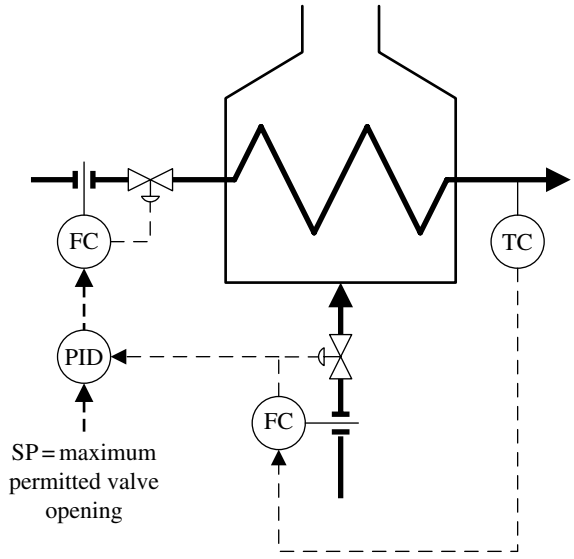

Figure 8.2 *Example of SISO constraint controller*

On our case study heater, the process operator will typically enter a SP of around 90%. This effectively converts a hard constraint to a soft constraint. In order to maintain control of the outlet temperature during minor process disturbances some leeway is required. This means that the process capacity is not fully utilised. Conditioning the constraint, as described in Chapter 5, offers an alternative method of converting it to a soft constraint and would permit this leeway to be reduced.

Tuning the PID controller is carried out in much the same way as normal. First the fuel flow and outlet temperature controllers should be properly tuned since their tuning affects how the fuel control valve moves when feed rate is changed. Model identification is completed by stepping the SP of the feed flow controller and observing how the fuel flow controller output varies. The tuning method outlined in Chapter 3 is followed although it is usual to tune less aggressively than normal by using a PI controller, with the gain reduced to around 25% of the calculated value.

Figure 8.3 shows a slightly different use of the same technique. Because of the wide range over which we wish to control flow it is necessary to use two control valves, one larger than the other. It would be possible to use the split range technique, as described in Chapter 5, so that to increase the flow the larger valve begins opening once the smaller valve is fully open. However, this could result in the problems with controller tuning as described in Chapter 5. Instead a VPC (shown as PID on Figure 8.3) is used with the flow controller output as its PV with its SP typically set at 50%. This will keep the smaller valve in the optimum position to make small corrections to the flow as necessary. Tuning follows much the same procedure. With the flow controller in automatic and the VPC in manual, small step changes are made to the position of the larger valve and the behaviour of the flow controller output recorded. Alternatively closed loop testing might be preferred with steps made to the SP of the VPC controller. Model identification and the calculation of tuning constants are completed as normal. Again it is common to use a PI controller with a reduced gain.

Although only occasionally a problem, the VPC does disturb the process. It causes a change in flow which the FC must correct. To help break the interaction between the two controllers, the VPC can be

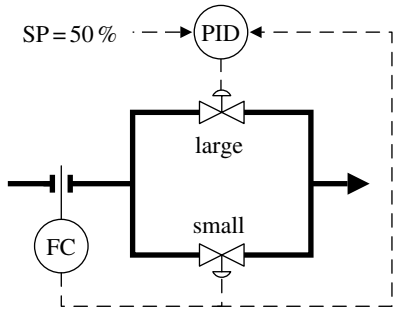

Figure 8.3 Flow control with two valves

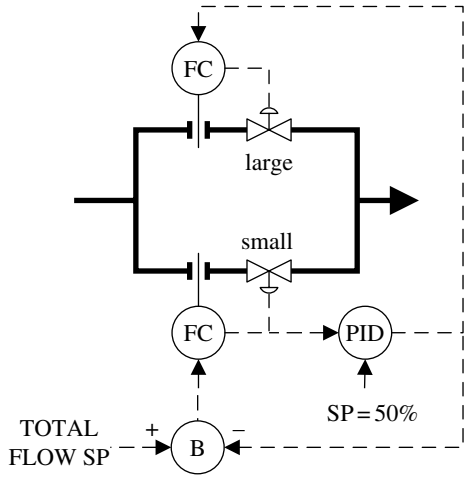

Figure 8.4 Flow control with two flow controllers

configured as a gap controller. For example setting a gap of ±10% around the SP would mean it would only take corrective action if the valve position is outside 40–60%. As the valve position is brought back into this range any interaction will cease. However, in order to achieve the required accuracy over a wide range of flows, it may also be necessary also to have two flow meters. This opens up the possibility of installing two separate flow controllers. This then permits the use of a simple feedforward scheme to break the interaction. Figure 8.4 shows how the bias algorithm, described in Chapter 6, may be used so that corrections made by the constraint controller do not change the total flow.

8.3 Signal Selectors

Signal selectors are used in the most basic form of multivariable control, i.e. multi-input single-output (MISO) applications. If in the previous example we instead wish to maximise the flow of one of two alternative supplies, for example to a hydraulic constraint, then the scheme shown as Figure 8.5 might be

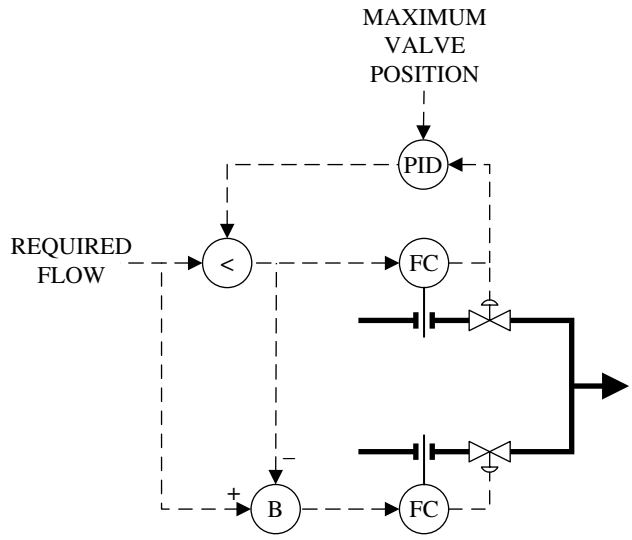

Figure 8.5 *Giving preference to one of two flow controllers*

applied. This approach is sometimes known as *auctioneering*. The SP of the required flow is overridden by the low signal selector (shown as < in Figure 8.5) if it causes the preferred flow controller output to exceed the maximum valve position. A simple bias feedforward scheme has been included to adjust the SP of the less preferred flow controller to make up the shortfall.

The fired heater described in previous chapters has, in addition to the maximum fuel valve position constraint, a maximum limit on burner pressure. This is approached by increasing feed rate. In order to operate at maximum feed rate the controller must be able to continuously identify which is the more limiting constraint. Figure 8.6 illustrates one possible configuration. Instead of the error *(PV – SP)* being calculated in the controller, it is calculated in each of the two bias algorithms. The two errors are compared in the signal selector and the selected value becomes the PV of the PID controller. The controller SP is fixed at 0 and so it will manipulate the SP of the feed flow controller until the first of the process limitations is reached.

We choose whether to use a low or high signal selector based on what would make sense to the process operator. In this case the feed rate would be determined by the constraint requiring the lower value, and so a low selector would seem preferable. However, because of our definition of error and that the constraints are upper limits, the scheme would normally require a high signal selector that would pass through the larger decrease in feed. The PID controller would then be configured as reverse-acting. Instead we include a negative gain in the bias algorithms; this allows the use of a low signal selector and the controller is then configured as direct-acting.

While simple in principle, controller tuning requires special attention. While it is clear which is the more limiting constraint if one is being violated and the other not, it is not so clear if both show that there is spare capacity. Selection needs to be based on which constraint will be reached first as feed rate is increased. It is tempting to use the process gains to assist this selection. If the process gain between fuel valve position and feed rate is $(K_p)_v$ and that between burner pressure and feed rate is $(K_p)_p$, then the gain term included in the valve bias algorithm is given by $-1/(K_p)_v$ and that in the burner pressure bias algorithm is $-1/(K_p)_p$. The output of each bias algorithm is then the permitted increase in feed rate. Choosing the lower of these two values would result in controlling against the more limiting constraint.

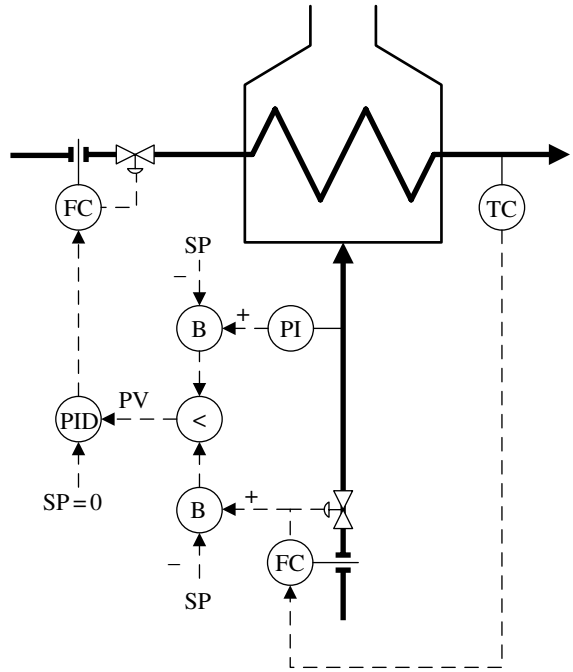

Figure 8.6 *Example of MISO constraint controller using single PID*

However, we also have to consider the tuning of the PID controller. If the dynamics of each constraint are different then the controller will require different tuning depending on which constraint is selected. A better approach is to move the controller gain from the PID to the biases. Process dynamics are obtained as usual by step-testing the feed flow SP. The dynamics of the fuel valve position are used to develop a set of tuning constants and the resulting controller gain (with its sign changed) entered as the gain in its bias algorithm. The dynamics of the burner pressure are used to develop a second set of tuning constants. The controller gain (with its sign changed) is used as the gain in its bias algorithm. The controller gain in the direct-acting PID algorithm can now be set to 1, although a lower value may be used as in the SISO example.

As with previous examples, derivative action is usually excluded. The integral time used in the controller is the average of the two determined for each constraint. If the values are far apart this might result in compromise tuning that gives poor response no matter which constraint is active. An approach which resolves this problem is shown in Figure 8.7. In this configuration each constraint has its own PID controller – a PC on burner inlet and a VPC on the fuel valve. In this example both constraint controllers are configured as reverse-acting. The signal selector must now include anti-reset windup to prevent the output of the unselected controller from saturating. In some DCS this is achieved by the use of the incremental version of the PID algorithm and so selection is based on the change in output of each controller. In others, using the full position version, this is achieved by setting the output of the unselected output to that of the one selected. Known as *reset feedback* this technique means that, at the next scan interval, selection is also based on changes in output. Measurement noise may now create a problem. The change in controller output caused by noise can readily exceed the change made to correct a deviation from SP.

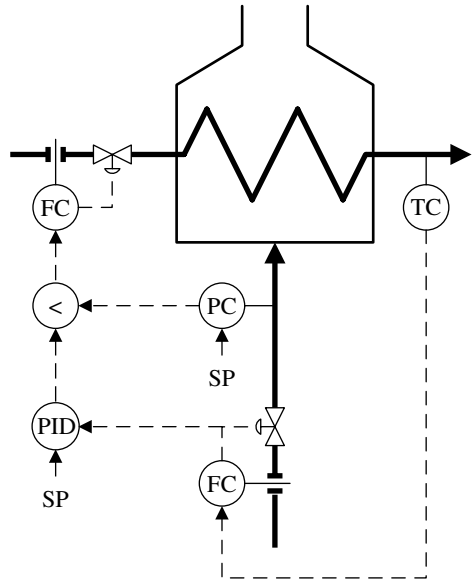

Figure 8.7 *Example of MISO constraint controller using multiple PIDs*

Imagine that the process is operating exactly at the fuel control valve limit and the burner pressure is well away from its limit. Noise peaks in the pressure measurement will be treated by the pressure controller as increases in pressure and will generate a negative change in controller output. This will be less than the zero change in output from the valve position controller and so will pass through the signal selector and cause a reduction in feed rate. Noise troughs will have no effect so, on average, the heater will then operate at a feed rate lower than that necessary to satisfy the constraints. In some situations this effect can be substantial, driving the feed rate well below where the operator would have set it manually.

Some DCS attempt to overcome this problem by including the option of an *override offset*. This offset is determined as the error (as a fraction of instrument range) on the unselected controller multiplied by its controller gain. If the controller is configured as reverse acting then the result of this calculation will be negative. For a low signal selector the output of the unselected controller is set to that of the selected controller plus the offset. For a high signal selector it is set to the output of the selected controller less the offset. A change in the output of the unselected controller that is smaller than the offset, such as might be generated by noise, will not now affect selection. While this does not guarantee that noise will not cause incorrect selection, it does reduce the probability. However, this also means that selection of a genuine violation is delayed until the SP is actually violated. This can be dealt with by setting a slightly more conservative constraint SP.

Filtering the pressure measurement might offer an alternative solution but the additional lag introduced may require noticeably slower controller tuning and therefore a substantially more conservative SP. If noise is not a problem in terms of excessive valve movement, the filtered measurement can be used to drive the signal selector and the unfiltered measurement used for control. One would need to check that the lag introduced by the filter does not unacceptably delay selection.

While it is possible to include additional constraints and signal selection, this type of strategy realistically is restricted to single output controllers. If, for example, it was permitted also to adjust the outlet temperature controller SP in order to approach the capacity limits, the resulting MIMO controller would

be extremely complex. Since both constraints would be affected by both MVs, simple selection logic cannot be applied. Decoupling might also be required to prevent the controllers from fighting each other. While theoretically a DCS block-based scheme could be designed, it is not an approach recommended.

8.4 Relative Gain Analysis

Relative gain analysis was developed [25] to assist with 'pairing' each PV with a MV and to assess the level of interaction. Relative gain is defined as the ratio of the process gain with all other controllers on manual to the same process gain with all other controllers in automatic mode. Ideally, placing other controllers in automatic should not affect the process gain of the first and so the relative gain would be 1.

For simplicity, consider first a 2 × 2 system. Plant testing, with all controllers on manual, has determined the process gains as follows:

	MV_1	MV_2
PV_1	$(K_p)_{11}$	$(K_p)_{12}$
PV_2	$(K_p)_{21}$	$(K_p)_{22}$

The *relative gain array (RGA)* is

	MV_1	MV_2
PV_1	λ_{11}	λ_{12}
PV_2	λ_{21}	λ_{22}

where λ is defined by

$$\lambda_{11} = \frac{\left(\frac{\Delta PV_1}{\Delta MV_1}\right)_{\Delta MV_2 = 0}}{\left(\frac{\Delta PV_1}{\Delta MV_1}\right)_{\Delta PV_2 = 0}} \tag{8.1}$$

While we have, from step-testing, the numerator (the process gain with all other controllers on manual) we cannot use step-testing to determine the denominator (the process gain with all other controller on auto) since we have yet to design the controllers. However, the process can be described by

$$\Delta PV_1 = \left(K_p\right)_{11} \Delta MV_1 + \left(K_p\right)_{12} \Delta MV_2 \tag{8.2}$$

$$\Delta PV_2 = \left(K_p\right)_{21} \Delta MV_1 + \left(K_p\right)_{22} \Delta MV_2 \tag{8.3}$$

From Equation (8.2), if $\Delta MV_2 = 0$, then

$$\left(\frac{\Delta PV_1}{\Delta MV_1}\right)_{\Delta MV_2 = 0} = \left(K_p\right)_{11} \tag{8.4}$$

From Equation (8.3), if $\Delta PV_2 = 0$, then

$$\Delta MV_2 = -\frac{\left(K_p\right)_{21}}{\left(K_p\right)_{22}} \Delta MV_1 \tag{8.5}$$

Substituting this into Equation (8.2)

$$\left(\frac{\Delta PV_1}{\Delta MV_1}\right)_{\Delta PV_2 = 0} = \left(K_p\right)_{11} - \frac{\left(K_p\right)_{12}\left(K_p\right)_{21}}{\left(K_p\right)_{22}} \tag{8.6}$$

Substituting Equations (8.4) and (8.6) into Equation (8.1)

$$\lambda_{11} = \frac{\left(K_p\right)_{11}\left(K_p\right)_{22}}{\left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21}} = \frac{1}{1 - \dfrac{\left(K_p\right)_{12}\left(K_p\right)_{21}}{\left(K_p\right)_{11}\left(K_p\right)_{22}}} \tag{8.7}$$

Provided the RGA is square, each row and each column sum to 1; so

$$\lambda_{12} = \lambda_{21} = 1 - \lambda_{11} \qquad \text{and} \qquad \lambda_{22} = \lambda_{11} \tag{8.8}$$

No matter which units are used for process gains, provided they are consistent, the relative gains are dimensionless.

There is an alternative method for determining λ_{11}. From Equations (8.2) and (8.3) if we step MV_1 while keeping MV_2 constant

$$\left(\Delta PV_1\right)_{\Delta MV_2 = 0} = \left(K_p\right)_{11} \Delta MV_1 \quad \text{and} \quad \left(\Delta PV_2\right)_{\Delta MV_2 = 0} = \left(K_p\right)_{21} \Delta MV_1 \tag{8.9}$$

$$\therefore \frac{\left(K_p\right)_{21}}{\left(K_p\right)_{11}} = \frac{\left(\Delta PV_2\right)_{\Delta MV_2 = 0}}{\left(\Delta PV_1\right)_{\Delta MV_2 = 0}} \tag{8.10}$$

Similarly, if we step MV_2 while keeping MV_1 constant

$$\left(\Delta PV_1\right)_{\Delta MV_1 = 0} = \left(K_p\right)_{12} \Delta MV_2 \quad \text{and} \quad \left(\Delta PV_2\right)_{\Delta MV_1 = 0} = \left(K_p\right)_{22} \Delta MV_2 \tag{8.11}$$

$$\therefore \frac{\left(K_p\right)_{12}}{\left(K_p\right)_{22}} = \frac{\left(\Delta PV_1\right)_{\Delta MV_1 = 0}}{\left(\Delta PV_2\right)_{\Delta MV_1 = 0}} \tag{8.12}$$

Substituting Equations (8.10) and (8.12) into Equation (8.7)

$$\lambda_{11} = \frac{\left(\frac{\Delta PV_1}{\Delta PV_2}\right)_{\Delta MV_2=0}}{\left(\frac{\Delta PV_1}{\Delta PV_2}\right)_{\Delta MV_2=0} - \left(\frac{\Delta PV_1}{\Delta PV_2}\right)_{\Delta MV_1=0}} \tag{8.13}$$

This provides a shortcut method. By first stepping MV_1 at constant MV_2 and determining the ratio between ΔPV_1 and ΔPV_2, then repeating the exercise by stepping MV_2, λ_{11} can then be calculated.

A value of 1 for λ_{11}, as we can see from Equation (8.1), means that the process gain is the same whether the other MV is changing or not. Thus, if there are no interactions, the RGA will be the *identity matrix*.

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{8.14}$$

This is clearly the ideal situation since it tells us that PV_1 can be controlled by MV_1 and PV_2 by MV_2 with no interaction between the controllers. Alternatively we may get the result

$$\Lambda = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{8.15}$$

This again tells us that there is no interaction but that PV_1 must be controlled by MV_2 and PV_2 by MV_1 – i.e. the pairing should be reversed. For λ_{11} to be 1 or 0, at least one element of the process gain matrix must be zero. If only one process gain is zero the controllers will display a one-way interaction. Two diagonal elements must be zero for there to be no interaction in either direction.

The worst possible case is

$$\Lambda = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} \tag{8.16}$$

A negative relative gain indicates that the process gain changes sign depending on whether the other controller is in automatic or manual mode. This will occur either because λ_{11} is itself less than zero or if it is greater than 1.

Control is impossible if the PVs are *parallel*. We can demonstrate this by first combining Equations (8.2) and (8.3) to eliminate ΔMV_1

$$\Delta MV_1 = \frac{\Delta PV_1 - \left(K_p\right)_{12} \Delta MV_2}{\left(K_p\right)_{11}} = \frac{\Delta PV_2 - \left(K_p\right)_{22} \Delta MV_2}{\left(K_p\right)_{21}} \tag{8.17}$$

Rearranging

$$\Delta PV_2 = \frac{\left(K_p\right)_{21} \Delta PV_1 + \Delta MV_2 \left(\left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21}\right)}{\left(K_p\right)_{11}} \tag{8.18}$$

Parallel PVs arise if a column in the process gain matrix is an exact multiple (m) of another.

$$\left(K_p\right)_{12} = m\left(K_p\right)_{11} \quad \text{and} \quad \left(K_p\right)_{22} = m\left(K_p\right)_{21} \tag{8.19}$$

Substituting in Equation (8.18)

$$\Delta PV_2 = \frac{\left(K_p\right)_{21}}{\left(K_p\right)_{11}} \Delta PV_1 \tag{8.20}$$

This means that if one PV is changed then the other must also change. For example, if PV_1 is at its SP and PV_2 is not, we want ΔPV_1 to be zero and we cannot therefore correct PV_2. This can be detected by considering the determinant $|\mathbf{K}|$ of the process gain matrix $\mathbf{K}$.

$$|\mathbf{K}| = \left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21} \tag{8.21}$$

This is sometimes described as the *controllability*. Substituting Equation (8.19) into Equation (8.21) shows that the determinant of the matrix will be zero if one column is an exact multiple of the other. Similarly |K| will be zero if one row is an exact multiple (n) of another.

$$\left(K_p\right)_{21} = n\left(K_p\right)_{11} \quad \text{and} \quad \left(K_p\right)_{22} = n\left(K_p\right)_{12} \tag{8.22}$$

If the determinant is zero then the relative gain is infinite. We can see this by combining Equation (8.7) with Equation (8.21)

$$\lambda_{11} = \frac{\left(K_p\right)_{11}\left(K_p\right)_{22}}{|\mathbf{K}|} \tag{8.23}$$

While relative gain calculations might indicate a possible control structure, care must be taken in the robustness of the conclusion. Process gains are unlikely to remain exactly constant and small changes can have a major impact on the feasibility of a design. For example, a process which we will consider further in a later chapter has the process gain matrix

$$\mathbf{K} = \begin{pmatrix} -0.962 & 4.17 \\ 0.806 & -5.32 \end{pmatrix} \tag{8.24}$$

From Equations (8.7) and (8.8) we find that

$$\Lambda = \begin{pmatrix} 2.91 & -1.91 \\ -1.91 & 2.91 \end{pmatrix} \tag{8.25}$$

However, Figure 8.8 shows the impact on this result if, for example, the value of $(K_p)_{12}$ were to change. While a reduction would reduce the level of interaction, only a relatively small increase would cause severe problems. From Equation (8.21) we can deduce that $|\mathbf{K}|$ will be zero when $(K_p)_{12}$ reaches 6.35

(as also shown by Figure 8.9). Above this value, as Figure 8.8 shows, the pairing should be reversed. Thus a controller, that might be stable under the conditions at which the process gains were determined, could become extremely unstable as conditions vary.

Relative gain analysis, in its simplest form, overlooks another potential problem. If we combine Equations (8.2) and (8.3) to eliminate ΔMV_2, we obtain

$$\Delta MV_1 = \frac{\left(K_p\right)_{22}\Delta PV_1 - \left(K_p\right)_{12}\Delta PV_2}{\left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21}} \tag{8.26}$$

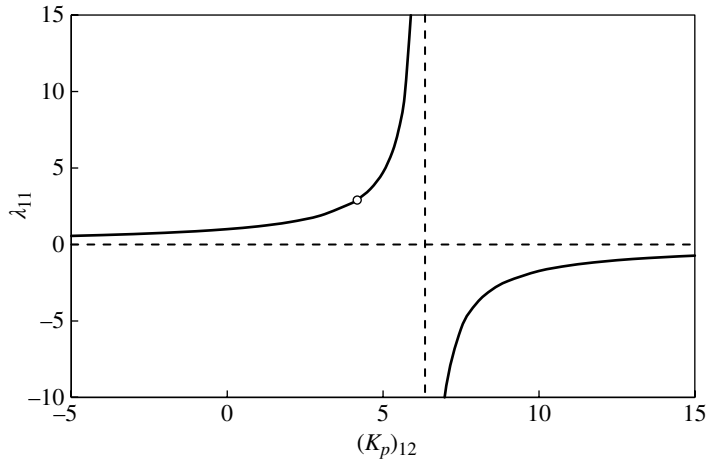

Figure 8.8 *Effect of process gain on relative gain*

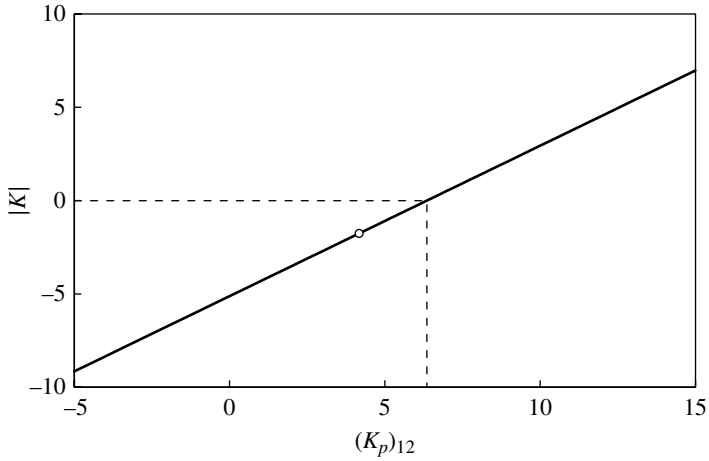

Figure 8.9 *Effect of process gain on determinant of gain matrix*

Dividing ΔPV_1 by this equation gives

$$\frac{\Delta PV_1}{\Delta MV_1} = \frac{\left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21}}{\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(\Delta PV_2 / \Delta PV_1\right)} \tag{8.27}$$

This, by definition, is the process gain 'seen' by the first of our controllers. It is, however, not a constant but dependent on the size of any change made to PV_2 – more specifically the ratio of this change to that in PV_1. In other words, should the SP of both controllers be changed then the process gain will vary depending on the size of the changes made. This is illustrated by Figure 8.10. Of greatest concern is that the sign of the process gain can change. This occurs at the point where the denominator of Equation (8.27) becomes zero, i.e.

$$\frac{\Delta PV_2}{\Delta PV_1} = \frac{\left(K_p\right)_{22}}{\left(K_p\right)_{12}} \tag{8.28}$$

In our example this evaluates to –1.28. For example, Figures 8.11 and 8.12 show the performance of the two controllers tuned using the method recommended in Chapter 3. As can be seen, ΔSP_1 was 2.0 and ΔSP_2 was –2.6 thus giving a ratio of –1.3. Figures 8.13 and 8.14 show the impact (with tuning suitably adjusted) of instead slightly reducing ΔSP_2 to –2.4 – now giving a ratio of –1.2 (the other side of the critical value). For the same change in PV_1, comparison of the MV_1 responses in Figures 8.12 and 8.14 shows that in the first case there was a steady state decrease in its value, while in the second case it increased. Clearly the two cases required substantially different controller tuning. Or, since it is not practical to automatically modify the tuning, retaining the same tuning would result in the controller frequently becoming unstable.

Figure 8.10 shows this problem does not arise if $(K_p)_{12}$ is zero and therefore λ_{11} is 1. As Equation (8.27) shows the effective process gain (shown by the coloured line) remains constant as $(K_p)_{11}$. However, this

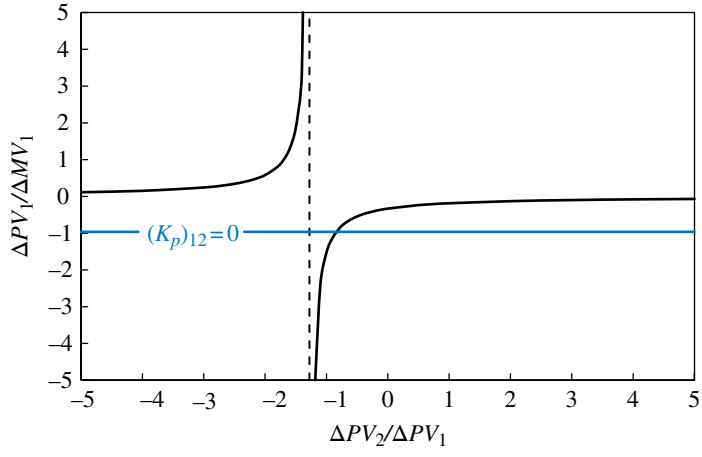

Figure 8.10 *Effect of disturbance on process gain between PV_1 and MV_1*

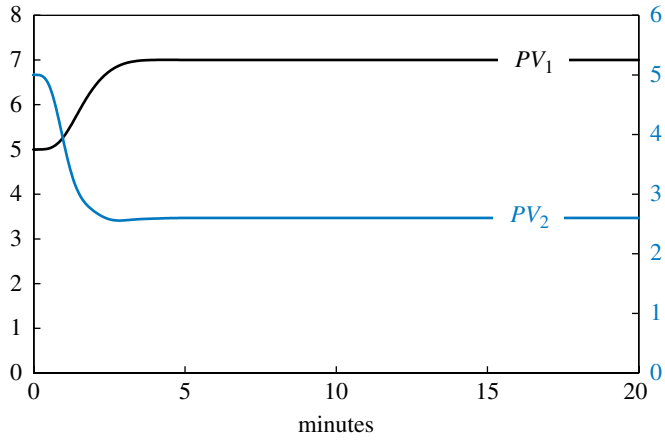

Figure 8.11 *Change in PV ($\Delta PV_1/\Delta PV_2 < -1.28$)*

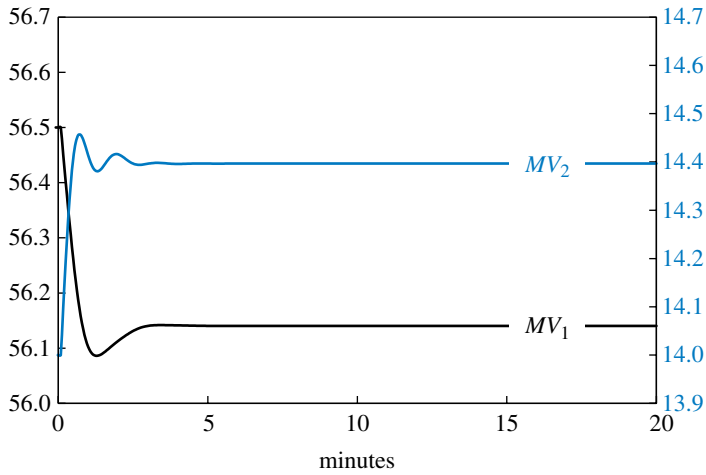

Figure 8.12 *Change in MV ($\Delta PV_1/\Delta PV_2 < -1.28$)*

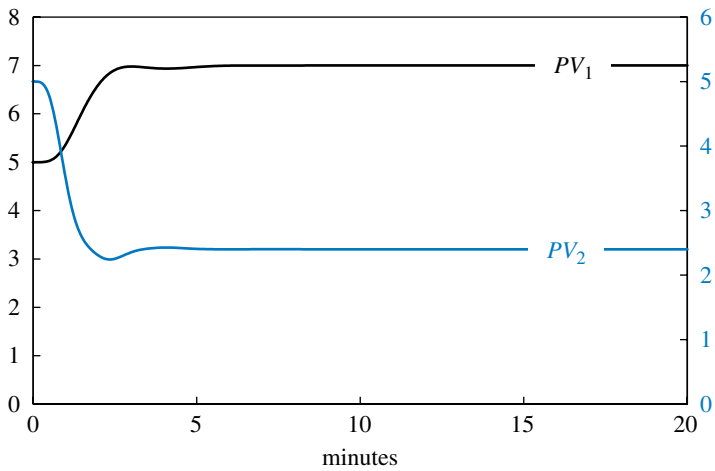

Figure 8.13 *Changes in PV ($\Delta PV_1/\Delta PV_2 > -1.28$)*

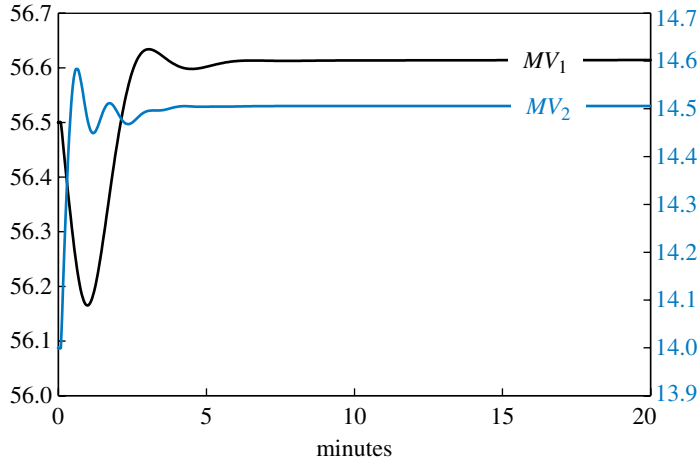

Figure 8.14 *Change in MV ($\Delta PV_1/\Delta PV_2 > -1.28$)*

does not eliminate the problem with the other controller. By taking a similar approach, first eliminating ΔMV_1 from Equations (8.2) and (8.3), then dividing ΔPV_2 by the result, we obtain

$$\frac{\Delta PV_2}{\Delta MV_2} = \frac{\left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21}}{\left(K_p\right)_{11} - \left(K_p\right)_{21}\left(\Delta PV_1 / \Delta PV_2\right)} \tag{8.29}$$

As can be seen from Equation (8.7), λ_{11} will be 1 if either $(K_p)_{12}$ or $(K_p)_{21}$ is zero. As Figure 8.15 shows, if $(K_p)_{12}$ is zero but $(K_p)_{21}$ is not, the process gain between PV_2 and MV_2 can still change sign. In other words it is not sufficient for the relative gain matrix to be the identity matrix; both $(K_p)_{12}$ and $(K_p)_{21}$ must be zero.

While the RGA for a 2×2 system gives some information about interactions, it usually helps little with pairing other than confirm the decision already made by the engineer. However, for larger problems, pairing is not always so obvious. For such problems we apply matrix techniques. Consider the process gain matrix for an $n \times n$ problem:

$$\mathbf{K} = \begin{pmatrix} \left(K_p\right)_{11} & \left(K_p\right)_{12} & . & \left(K_p\right)_{1n} \\ \left(K_p\right)_{21} & \left(K_p\right)_{22} & . & \left(K_p\right)_{2n} \\ . & . & . & . \\ \left(K_p\right)_{n1} & \left(K_p\right)_{n2} & . & \left(K_p\right)_{nn} \end{pmatrix}. \tag{8.30}$$

It can be shown that the relative gain matrix can be determined by first inverting the process gain matrix to give $\mathbf{K}^{-1}$, transposing it to give $(\mathbf{K}^{-1})^{\mathrm{T}}$ and multiplying by $\mathbf{K}$.

$$\Lambda = \left(\mathbf{K}^{-1}\right)^T * \mathbf{K} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & . & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & . & \lambda_{2n} \\ . & . & . & . \\ \lambda_{n1} & \lambda_{n2} & . & \lambda_{nn} \end{pmatrix} \tag{8.31}$$

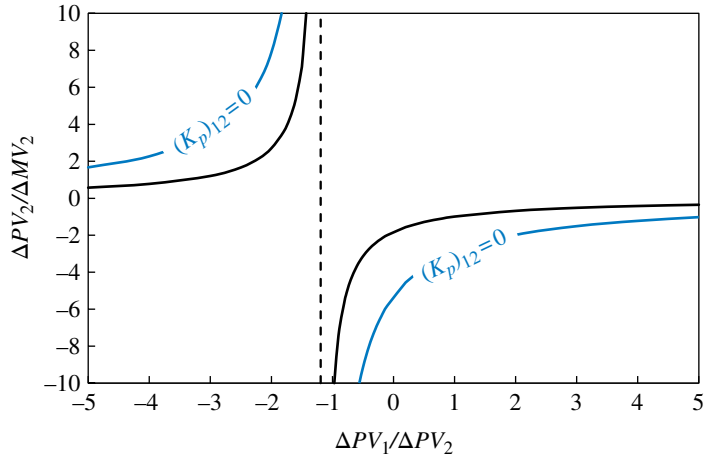

Figure 8.15 *Effect of disturbance on process gain between PV_2 and MV_2*

Note that the operator * is not the conventional row-by-column matrix multiplication but an element-by-element multiplication – resulting in what is known as the *Hadamard* (or *Schur*) *product*.

It is common in MPC that there are more control variables (CVs) than manipulated variables (MVs) and so the gain matrix will not be square and therefore cannot be inverted normally. Under these circumstances the *pseudo-inverse* can be calculated as follows:

$$\mathbf{K}^{-1} = \left(\mathbf{K}^T.\mathbf{K}\right)^{-1}.\mathbf{K}^T \tag{8.32}$$

As with the square RGA, the sum of relative gains in each column will sum to 1 but this is no longer the case for each row. In the unlikely event that the matrix has more columns than rows, i.e. there are more MVs than CVs, then the rows will sum to 1 but the columns will not and a slightly different calculation is required.

$$\mathbf{K}^{-1} = \mathbf{K}^T.\left(\mathbf{K}.\mathbf{K}^T\right)^{-1} \tag{8.33}$$

While these calculations may look complex they are readily configurable in a spreadsheet package – most of which support matrix functions. Alternatively there are a number of commercially available packages which will also analyse the RGA and suggest control strategies.

When using the RGA to decide pairing, any PV/MV combination where λ is negative should be avoided. This means that the process gain for the proposed controller reverses sign as other controllers are switched between auto and manual – leading immediately to the controller saturating. The procedure is to next check if any row or column has only one positive element. If so, then this decides the first pairing. If not the case then the λ in the matrix that is closest to 1 is chosen to provide the first pairing. All other elements in the same column or row as this element are then ignored and the process repeated until all pairings are complete. It is quite possible that not all variables may be paired.

With the advent of MPC there are mixed views about the value of relative gain analysis. However, it can be useful in checking the design of MPC applications. Large MPC strategies are difficult for the process operator to understand and difficult to maintain. The RGA can be used to identify for consideration any controllers that can be moved outside MPC as SISO controllers. It can also be used to see if MPC can be broken into separate smaller controllers. However, relative gain analysis takes no account of process dynamics. Variables can show no interaction at steady state but during disturbances their transient behaviour can cause control problems. Further, the analysis assumes linear behaviour. As operating conditions change it is likely that some of the process gains will vary. As we have seen, recalculating the RGA under these conditions may alter the conclusions in terms of both pairing and decoupling. For these reasons the RGA should only be considered as a technique for eliminating some of the pairing options and suggesting configurations that need to be evaluated further.

8.5 Niederlinski Index

Strongly related to relative gain, the *Niederlinski Index (NI)* is another method of checking pairing. It is calculated from the determinant of the process gain matrix and the product of all the process gains on the diagonal of the matrix.

$$NI = \frac{|\mathbf{K}|}{\prod_{i=1}^{n} K_{ii}} \tag{8.34}$$

Should this index be infinite then the process gain of at least one pairing is zero and so control would be impossible. Ideally the value should be unity, showing that there is no interaction. A value of zero indicates, as described above, that there are parallel PVs. Should the index be negative then the proposed control configuration will definitely be unstable – no matter what the tuning.

For example the calculation below, for the matrix described by Equation (8.24), would suggest that there is significant interaction.

$$NI = \frac{(-0.962 \times -5.32) - (4.17 \times 0.806)}{(-0.962 \times -5.32)} = 0.343 \tag{8.35}$$

Figure 8.16 shows the effect of varying $(K_p)_{12}$ in the process gain matrix described by Equation (8.24). As its value approaches 6.35, *NI* approaches zero indicating that the PVs are then parallel. Increasing the gain beyond this point causes the *NI* to become negative, indicating that the pairing should be reversed.

Note that only for the 2 × 2 case is *NI* equal to the reciprocal of λ_{11}. Further, a positive value only guarantees stability in the 2 × 2 case. For larger problems the index can only be used to reject schemes proven to be unstable.

The technique is only applicable to a square matrix and so therefore cannot normally be applied to MPC.

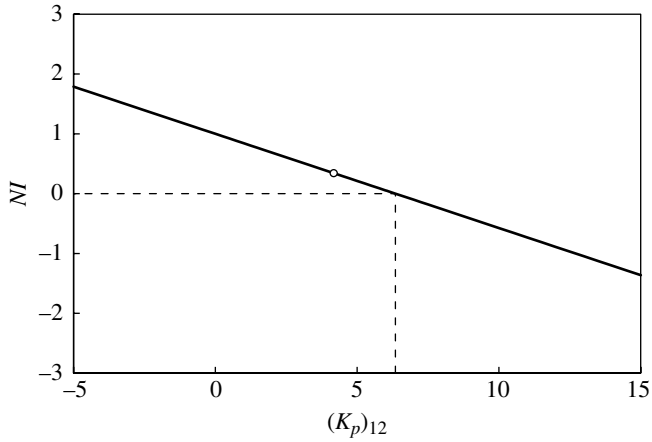

Figure 8.16 *Effect of process gain on Niederlinski Index*

8.6 Condition Number

The *condition number* is more commonly used to assess the sensitivity of the solution, of a set of simultaneous equations described by **A.x = y**, to small changes in the values of **y**. A *well-conditioned* matrix will have a condition number close to 1. The condition number will be infinite if the matrix is singular. It is this property that makes the technique of interest to the control engineer. It indicates that the matrix contains parallel PVs. There are mixed views in industry about what is a realistically maximum acceptable value for the condition number. Some would describe the matrix as *ill-conditioned* and modify the controller if the value exceeds 5; others have demonstrated effective control where the value is in excess of 15.

The calculation of condition number is usually published using symbols and terminology with which many control engineers will be unfamiliar. However, the calculation is quite simple. The condition number, $\kappa_1(\Lambda)$, is derived from the *norm* of the relative gain matrix and its inverse. The norm is defined as the maximum absolute column sum; we take the absolute value of all the relative gains in the matrix, sum each column and select the largest result.

$$\left\|\Lambda\right\|_1 = \max_j \sum_{i=1}^{n} \left|\lambda_{ij}\right| \tag{8.36}$$

We do the same for the inverse of the matrix and so calculate the condition number.

$$\kappa_1\left(\Lambda\right) = \left\|\Lambda_1\right\|.\left\|\Lambda^{-1}\right\|_1 \tag{8.37}$$

The condition number, $\kappa_\infty(\Lambda)$, can also be determined from the maximum absolute row sum.

$$\left\|\Lambda\right\|_\infty = \max_i \sum_{j=1}^{n} \left|\lambda_{ij}\right| \tag{8.38}$$

$$\kappa_\infty\left(\Lambda\right) = \left\|\Lambda_\infty\right\|.\left\|\Lambda^{-1}\right\|_\infty \tag{8.39}$$

For example the calculations below, for the relative gain matrix described by Equation (8.25), show that it may be sufficiently well-conditioned not to cause a control problem.

$$\left\|\Lambda\right\|_1 = \left\|\Lambda_\infty\right\| = 4.82 \tag{8.40}$$

$$\Lambda^{-1} = \begin{pmatrix} 0.604 & 0.396 \\ 0.396 & 0.604 \end{pmatrix} \tag{8.41}$$

$$\left\|\Lambda^{-1}\right\|_1 = \left\|\Lambda^{-1}\right\|_\infty = 1 \tag{8.42}$$

$$\kappa_1\left(\Lambda\right) = \kappa_\infty\left(\Lambda\right) = 4.82 \tag{8.43}$$

Figure 8.17 shows the effect of varying $(K_p)_{12}$ in the process gain matrix described by Equation (8.24). As its value approaches 6.35, κ becomes infinite – indicating that the PVs are parallel. As $(K_p)_{12}$ approaches zero, κ approaches 1 – indicating that interaction has been broken between PV_1 and MV_2 (but not between PV_2 and MV_1).

Remembering that each row and column in a square relative gain matrix sum to 1 then, only if all the relative gains are positive, will the norm be 1. The same also applies to the norm of the inverse of the relative gain matrix (as illustrated in this example). For matrices larger than 3 × 3, κ_1 is unlikely to be exactly the same as κ_∞. The difference, however, will be sufficiently small not to affect any conclusion drawn from the result.

The formulae for calculating the condition number are applicable to non-square matrices. In addition to applying Equation (8.32), we can also calculate the RGA and its condition number for all possible 2 × 2 pairings. Consider the process gain matrix

$$\mathbf{K} = \begin{pmatrix} -42 & 5 & -13 \\ 10 & -45 & 5 \\ -7 & 34 & 9 \end{pmatrix} \tag{8.44}$$

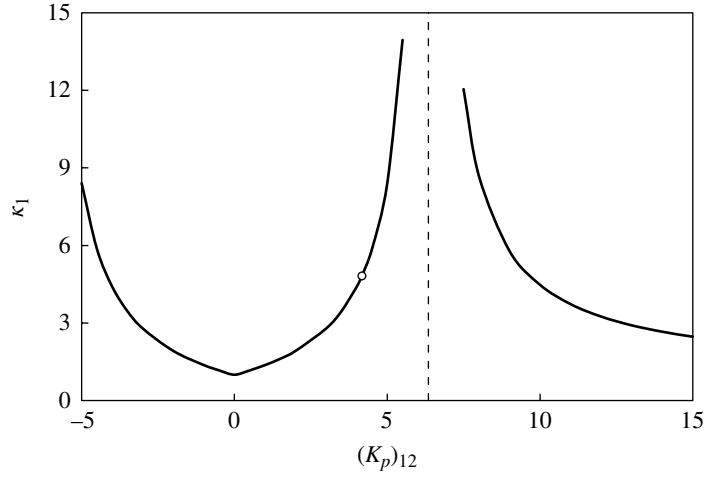

Figure 8.17 *Effect of process gain on condition number*

Applying Equation (8.31) gives:

$$\Lambda = \begin{pmatrix} 1.041 & -0.027 & -0.014 \\ -0.210 & 0.910 & 0.300 \\ 0.169 & 0.117 & 0.714 \end{pmatrix} \tag{8.45}$$

Examination of the relative gains would suggest that the process gain matrix is well conditioned. Indeed, applying Equations (8.36) and (8.37) shows that the condition number is 2.8. However, selecting one of the nine possible 2 × 2 sub-sets, gives

$$\mathbf{K} = \begin{pmatrix} 10 & -45 \\ -7 & 34 \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} 13.6 & -12.6 \\ -12.6 & 13.6 \end{pmatrix} \tag{8.46}$$

This has a condition number of 26.2 – indicating that PV_2 and PV_3 could not both be controlled using only MV_1 and MV_2. However, the rigorous analysis showed the addition of MV_3 permits both these and PV_1 to be controlled. We will describe later how MVs and PVs can be selectively *dropped* from MPC. If the operator were to drop MV_3, it would be necessary to also drop either PV_2 or PV_3.

8.7 Steady State Decoupling

Figure 8.18 shows an example of a highly interactive control problem. Two streams of different temperature are blended to meet a required temperature but also they must be manipulated to control the drum level. Changing either flow will change both the temperature and the level. If both controllers were independently tuned for tight control the likelihood is that, with both on auto, they would be unstable.

One possibility is that one of the controllers could be de-tuned so that it reacted very slowly to disturbances. For example, it might be acceptable for the level to deviate from SP for long periods and changes in its SP are likely to be rare. Priority would then be given to the temperature controller. While the interaction still exists, this would avoid instability.

We can apply relative gain analysis to determine the level of interaction. If F_1 and T_1 are the flow and temperature of one stream, and F_2 and T_2 are for the other, the combined temperature is given by

$$T = \frac{F_1 T_1 + F_2 T_2}{F_1 + F_2} \tag{8.47}$$

Differentiating to obtain the process gains

$$\left(K_p\right)_{11} = \frac{dT}{dF_1} = \frac{F_2\left(T_1 - T_2\right)}{\left(F_1 + F_2\right)^2} \tag{8.48}$$

$$\left(K_p\right)_{12} = \frac{dT}{dF_2} = \frac{F_1\left(T_2 - T_1\right)}{\left(F_1 + F_2\right)^2} \tag{8.49}$$

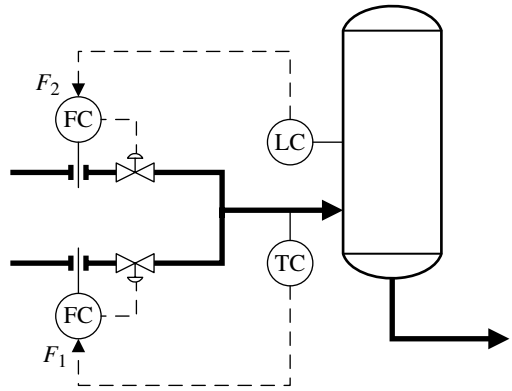

Figure 8.18 *Highly interactive controllers*

If the cross-sectional area of the drum is A, then (because it is an integrating process) we predict the rate of change of level, rather than the level itself.

$$\frac{dL}{dt} = \frac{\left(F_1 + F_2\right)}{A} \tag{8.50}$$

Again differentiating to obtain the process gains

$$\left(K_p\right)_{21} = \frac{d\left(\frac{dL}{dt}\right)}{dF_1} = \frac{1}{A} \tag{8.51}$$

$$\left(K_p\right)_{22} = \frac{d\left(\frac{dL}{dt}\right)}{dF_2} = \frac{1}{A} \tag{8.52}$$

And, from Equations (8.7) and (8.8)

$$\Lambda = \begin{pmatrix} \frac{F_2}{F_1 + F_2} & \frac{F_1}{F_1 + F_2} \\ \frac{F_1}{F_1 + F_2} & \frac{F_2}{F_1 + F_2} \end{pmatrix} \tag{8.53}$$

We can see from Equation (8.47), that if T_2 approaches T (and T_1 does not), F_1 will approach zero and λ_{11} will approach 1. F_2 can then be used to control level with little impact on temperature. Similarly, if T_1 approaches T, then F_2 should be used to control level.

If F_1 and F_2 are similar, i.e. because the average of the two stream temperatures is close to the target temperature, the relative gains all approach 0.5 and the process is likely to be uncontrollable using the chosen MVs.

From Equation (8.34)

$$NI = \frac{F_2(T_1 - T_2) - F_1(T_2 - T_1)}{F_2(T_1 - T_2)} = \frac{F_1 + F_2}{F_2} \tag{8.54}$$

NI is positive and so it is possible, with suitable tuning, for the controller to be stable. If T_1 is equal to T_2 then *NI* is indeterminate, showing that there are parallel PVs. Control would only be possible if both temperatures are equal to the required temperature.

Figure 8.19 shows a partial solution. As in Figure 8.18, the TC manipulates F_1, but as the flow changes this is fed forward via a bias algorithm so that F_2 is changed by the same amount in the opposite direction. Thus the drum level will remain unchanged. Interaction has been broken in one direction; $(K_p)_{21}$ is now zero, and so the RGA becomes the identity matrix. If the LC takes corrective action this will change the temperature but the partial decoupling would be sufficient to stop the interaction causing instability.

Figure 8.20 shows an alternative partial solution. In this case the LC shown in Figure 8.18 has been retained, manipulating F_2. For the reasons given in Chapter 6 (see Figure 6.4) the temperature controller now manipulates the ratio between F_1 and the total flow. This ensures that when the LC takes corrective action, the temperature remains unchanged. Thus $(K_p)_{12}$ is now zero. Corrections made by the TC will affect the vessel level and so the interaction has been broken in only one direction.

Figure 8.21 shows a fully decoupled controller. Rather than use the flow measurements in the decoupling calculations, the controller outputs (equivalent to the flow SPs) have been used. This is an option. It gives a slight dynamic advantage because the changes are fed forward sooner and are also noise-free. However, it does assume that the flow controllers can achieve their SPs. If either controller saturates, or is switched to manual, then the decoupling is likely to cause windup problems.

8.8 Dynamic Decoupling

Dynamic decoupling, first installed using DCS function blocks, is now largely provided by proprietary MPC packages. We here describe the DCS approach primarily to help understand the principles of decoupling using non-proprietary techniques. For reasons that will become apparent the MPC package approach, although initially more costly, is usually the better solution economically.

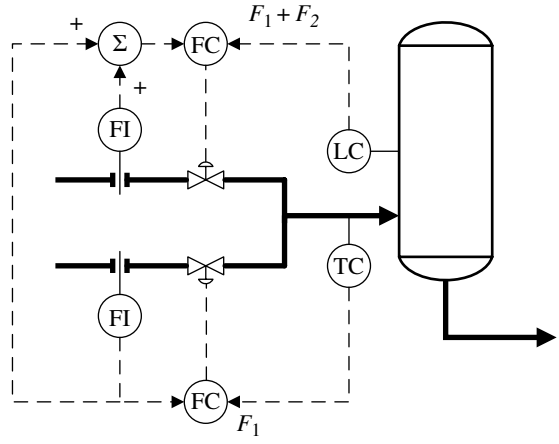

Figure 8.19 *Partially decoupled controller (using bias)*

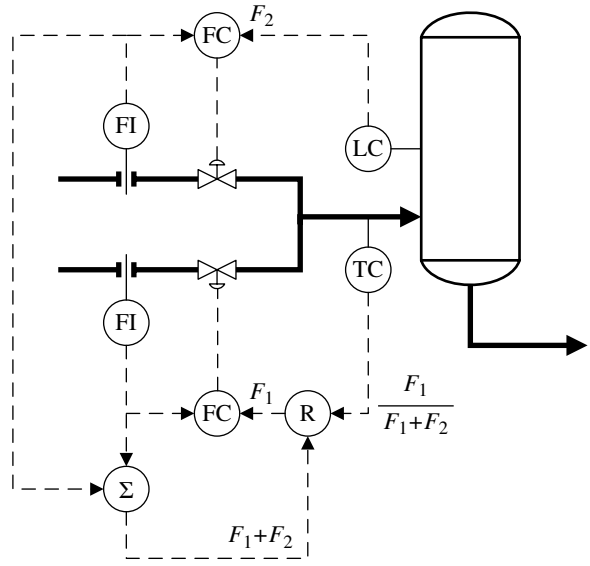

Figure 8.20 *Partially decoupled controller (using ratio)*

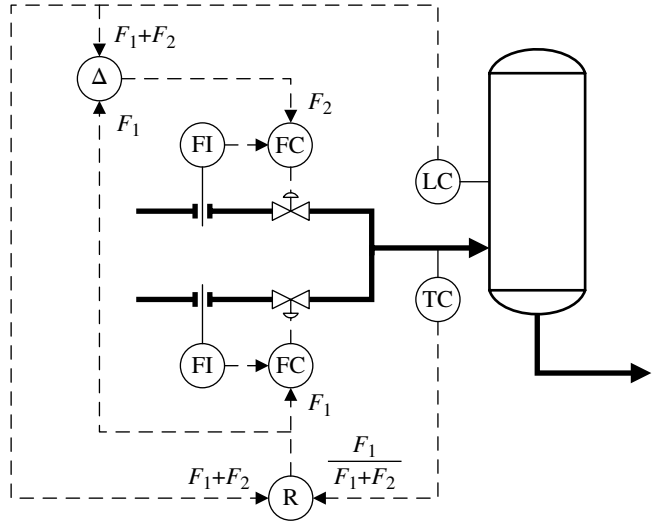

Figure 8.21 *Fully decoupled controller*

We will consider first a 2 × 2 system. Decoupling effectively feeds forward corrective action made by each controller to the MV of the other so that the other PV is undisturbed. In principle this results in two non-interacting controllers which can then be tuned conventionally. Figure 8.22 illustrates how both PVs are affected by both MVs. We assume we wish to operate where the contours of constant PV intersect. We start with PV_2 at SP_2, but PV_1 away from SP_1. The challenge is to move PV_1 to its SP without disturbing PV_2. To do so we have to make a compensating change to MV_2. Decoupling is tuned to determine the size

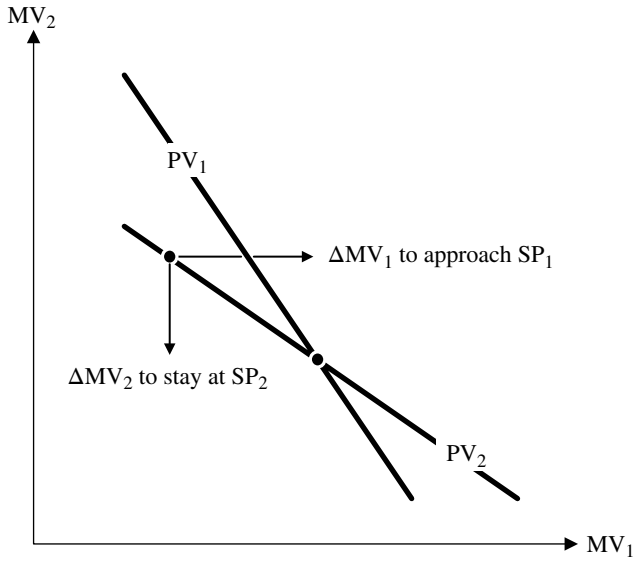

Figure 8.22 Decoupling principle

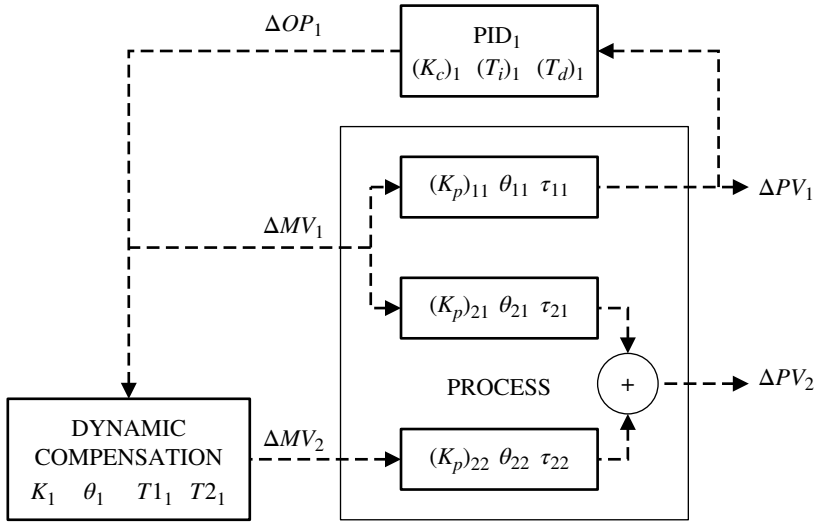

Figure 8.23 MV_1 to MV_2 decoupler

of this compensating move. If made at the same time as the adjustment to MV_1 then we will have achieved steady-state decoupling. Both PVs will ultimately reach their SPs but PV_2 is likely to show a transient deviation, i.e. it will start and finish on the PV_2 contour but not necessarily follow it. To ensure it does so, we also require dynamic compensation.

Figure 8.23 shows the first of the decouplers. When PID_1 takes corrective action, the decoupler applies dynamic compensation to the change in output (ΔOP_1) and makes a change to MV_2 that counteracts the

disturbance that the change in MV_1 would otherwise cause to PV_2. Dynamic compensation is provided by a deadtime/lead-lag algorithm.

The tuning method for the decoupler is exactly that described for bias feedforward in Chapter 6, i.e.

$$K_1 = -\frac{(K_p)_{21}}{(K_p)_{22}} \qquad \theta_1 = \theta_{21} - \theta_{22} \qquad T1_1 = \tau_{22} \qquad T2_1 = \tau_{21} \tag{8.55}$$

A second decoupler makes compensatory changes to MV_1 when PID_2 takes corrective action, as shown in Figure 8.24. Tuning for this second decoupler is given by

$$K_2 = -\frac{(K_p)_{12}}{(K_p)_{11}} \qquad \theta_2 = \theta_{12} - \theta_{11} \qquad T1_2 = \tau_{11} \qquad T2_2 = \tau_{12} \tag{8.56}$$

Figure 8.25 shows both decouplers, but without the PID controllers.

The addition of each decoupler has added another feedback path to each controller and thus changed the apparent process gain. In the example shown in colour, the output from PID1 (ΔOP_1) passes to the process as before as ΔMV_1 but also passes through the first dynamic compensation algorithm to become part of ΔMV_2. Since both ΔMV_1 and ΔMV_2 both affect ΔPV_1 the process gain changes to

$$\frac{\Delta PV_1}{\Delta OP_1} = (K_p)_{11} + K_1 (K_p)_{12} \tag{8.57}$$

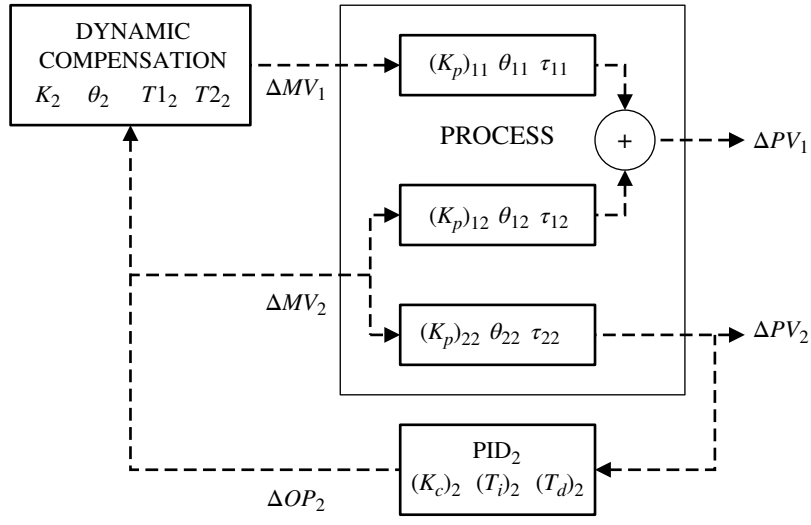

Figure 8.24 *MV_2 to MV_1 decoupler*

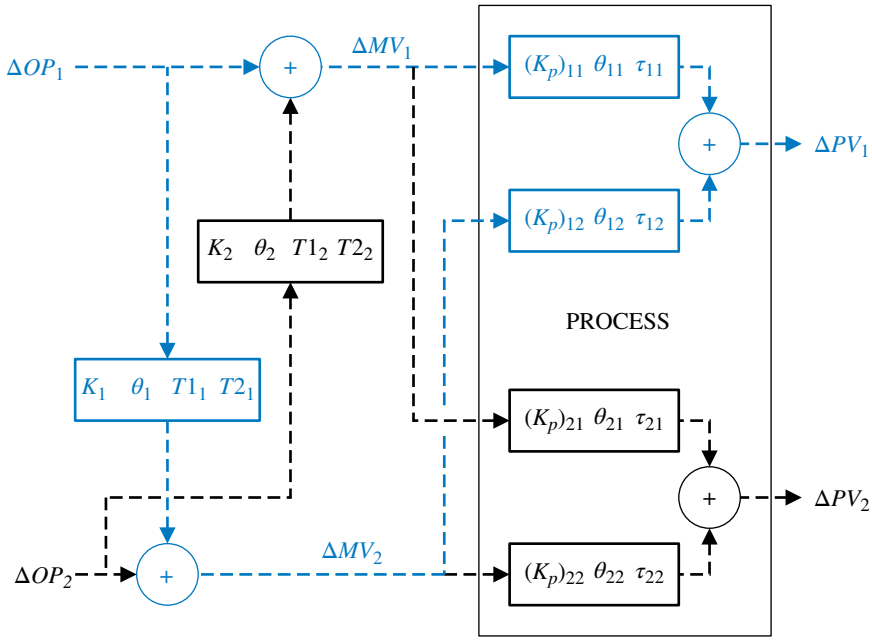

Figure 8.25 *2 × 2 dynamic decoupler*

Substituting for K_1 from Equation (8.55)

$$\frac{\Delta PV_1}{\Delta OP_1} = \left(K_p\right)_{11} - \frac{\left(K_p\right)_{12}\left(K_p\right)_{21}}{\left(K_p\right)_{22}} = \frac{\left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21}}{\left(K_p\right)_{22}} \tag{8.58}$$

By definition, the process gain without the decoupler in place is given by

$$\frac{\Delta PV_1}{\Delta OP_1} = \left(K_p\right)_{11} \tag{8.59}$$

Therefore, dividing Equation (8.59) by Equation (8.58), to maintain the loop gain constant the controller gain must be multiplied by

$$\frac{\left(K_p\right)_{11}\left(K_p\right)_{22}}{\left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21}} = \lambda_{11} \tag{8.60}$$

The same applies to the apparent process gain of PID_2. The process gain with the decoupler is

$$\frac{\Delta PV_2}{\Delta OP_2} = \left(K_p\right)_{22} + K_2\left(K_p\right)_{21} \tag{8.61}$$

Substituting from Equation (8.56) for K_2

$$\frac{\Delta PV_2}{\Delta OP_2} = \left(K_p\right)_{22} - \frac{\left(K_p\right)_{12}\left(K_p\right)_{21}}{\left(K_p\right)_{11}} = \frac{\left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21}}{\left(K_p\right)_{11}} \tag{8.62}$$

By definition, the process gain without the decoupler in place is given by

$$\frac{\Delta PV_2}{\Delta OP_2} = \left(K_p\right)_{22} \tag{8.63}$$

Therefore, dividing Equation (8.63) by Equation (8.62), to maintain the loop gain constant the controller gain must be multiplied by

$$\frac{\left(K_p\right)_{11}\left(K_p\right)_{22}}{\left(K_p\right)_{11}\left(K_p\right)_{22} - \left(K_p\right)_{12}\left(K_p\right)_{21}} = \lambda_{22} \tag{8.64}$$

If it is not the case that the decouplers are added to existing PID controllers, or the opportunity to optimally tune existing controllers is to be exploited, then these can be tuned according to the procedures described in Chapter 3. The process model to tune PID_1 is

$$K_p = \left(K_p\right)_{11} + K_1\left(K_p\right)_{12} \qquad \theta = \theta_{11} \qquad \tau = \tau_{11} \tag{8.65}$$

And that for PID_2 is

$$K_p = \left(K_p\right)_{22} + K_2\left(K_p\right)_{21} \qquad \theta = \theta_{22} \qquad \tau = \tau_{22} \tag{8.66}$$

Alternatively these models could be obtained by separately step-testing OP_1 and OP_2 with the decouplers in place.

As a partial solution to the problem of interaction, a one-way decoupler might be considered. So if, for example, it was more important to keep PV_1 rather than PV_2 close to SP, then decoupler 2 would be implemented without decoupler 1. This would be sufficient to prevent the interaction causing instability and would be considerably easier to implement and maintain. The disadvantage of course is that the control of PV_2 would be poor.

While the amount of step-testing required for the DCS approach is identical to that for a proprietary MPC package, and the tuning calculations very simple, implementation in the DCS is quite complex. Not only does it involve a large number of DCS blocks but a great deal of engineering must be put into properly scaling each block and ensuring bumpless transfer from manual to auto.

If, after commissioning, it was discovered that an equipment constraint was being frequently encountered then we would want to include this constraint in the controller. In the DCS case this is likely to greatly increase complexity and would require major re-engineering. With a MPC package the addition would be much simpler.

It is difficult in the DCS to make provision for one of the controllers to be disabled without the other(s). While this might be overcome by only permitting all-or-nothing operation, one controller may saturate. This will change the apparent process gain of the other(s). A properly configured MPC package will handle this situation routinely.

8.9 MPC Principles

It is not the intention here to reproduce the detailed theory of model predictive control (MPC). This has become an almost obligatory section in modern control texts. There are also numerous papers, marketing material and training courses. In this book its description is limited to its general principles; focus instead is on how to apply it and monitor its performance.

MPC packages differ from PID type controllers in a number of aspects. Firstly PID type controllers require a SP. MPC requires only HI and LO limits to be set for each *controlled variable* (CV). The HI limit can be set to the LO limit if a true SP is required. MPC will exploit the range between the HI and LO limits in two ways – either to avoid a violating a constraint elsewhere on the process or, if all constraints are satisfied, reduce the overall operating cost.

MPC also permits HI/LO constraints to be placed on *manipulated variables* (MV). While this is possible with PID controllers, this could cause a deviation from SP. MPC will use other MVs to satisfy a constraint if one has reached its limit.

Some MPC packages permit hard and soft constraints. Soft constraints are adhered to if possible but will be violated if this is the only way of avoiding violating a hard constraint. Other packages permit weighting to be applied to constraints so that the engineer can specify which should be violated first if MPC cannot identify a feasible solution.

Basic PID controls must wait for a disturbance to be measured before responding, whereas MPC predicts future deviations and takes corrective action to avoid future violation of constraints. There are two fundamentally different ways this is done. Some packages use high order Laplace transforms, others use a time series. A time series comprises a linear function of previous values of the MV (and sometimes also CV). In the function CV_n is the predicted next value of the CV; MV_{n-1} is the current value of the MV, MV_{n-2} the previous value, etc. The coefficients a_1, a_2 etc. are determined by regression analysis of step test data. Several such functions are in common use in proprietary MPC packages.

Finite in this context means until steady state; the *Finite Impulse Response (FIR)* for a SISO process is given by

$$CV_n = \sum_{i=1}^{n} a_i MV_{n-i} \qquad \text{or} \qquad \Delta CV_n = \sum_{i=1}^{n} a_i \Delta MV_{n-i} \tag{8.67}$$

The finite impulse response is the derivative, with respect to time, of the *Finite Step Response (FSR)* which is given by

$$CV_n = \sum_{i=1}^{n} a_i \Delta MV_{n-i} \tag{8.68}$$

Dynamic matrix control uses

$$CV_n = \sum_{i=1}^{n-1} a_i \Delta MV_{n-i} + a_n MV_{n-N} \tag{8.69}$$

Also used is

$$CV_n = \sum_{i=1}^{n} a_i MV_{n-i} + \sum_{i=1}^{n} b_i CV_{n-i} \tag{8.70}$$

These equations are then used to predict the effect of future changes in the MV. The *prediction horizon* is the number (M) of sample periods used in predicting the value of the CV. The *control horizon* is the number (N) of control moves (ΔMV) that are calculated into the future. In matrix form Equation (8.68), for example, therefore becomes

$$\begin{pmatrix} CV_1 \\ CV_2 \\ CV_3 \\ . \\ CV_M \end{pmatrix} = \begin{pmatrix} a_1 & 0 & 0 & . & 0 \\ a_2 & a_1 & 0 & . & 0 \\ a_3 & a_2 & a_1 & . & 0 \\ . & . & . & . & . \\ a_M & a_{M-1} & a_{M-2} & . & a_{M-N+1} \end{pmatrix} \begin{pmatrix} \Delta MV_0 \\ \Delta MV_1 \\ \Delta MV_2 \\ . \\ \Delta MV_N \end{pmatrix} \tag{8.71}$$

Future moves are calculated to minimise the sum of the squares of the predicted deviations from target over the control horizon. While the next N moves are calculated, only the first (ΔMV_0) is implemented. At the next cycle the controller recalculates a new set of control moves. This will account for any prediction errors or unmeasured load disturbances.

All MPC packages include at least one parameter that permits the engineer to establish the compromise between fast approach to target and MV movement. Such an approach was adopted in the PID tuning methods described in Chapter 3. Some MPC packages permit the engineer to define *move suppression* which penalises large changes to the MV. Others include a term similar to λ used in the Lambda tuning method and internal model control. This permits the engineer to define a trajectory for the approach to target.

While the deadtime compensation techniques covered in Chapter 7 provide similar functionality for SISO controllers these generally need to be custom configured in the DCS. MPC includes the feature as standard. Further the equations above predict the value of a CV based on the changes in a single MV. MPC sums the effect of all the MVs. Like other predictors the actual CV is compared to what was predicted and a bias term updated to compensate for any inaccuracy.

PID controllers are either in manual or automatic mode. MPC can be selectively switched to manual. Any MV can be *dropped*, in which case its SP can be changed by the process operator. When calculating future moves the controller effectively sets the HI/LO limits on this MV to its current SP. Some MVs may be categorised as *critical*, dropping one of these disables the whole MPC strategy. Any CV can similarly be dropped, in which case MPC ignores any constraints placed on it. Some MPC packages support the definition of sub-controllers. These allow part of MPC to be disabled while keeping the remainder in service. Even if disabled MPC continues to operate in *warm* mode. While it makes no change to any MV, it continues to update its CV predictions so that it operates correctly when re-commissioned. If prediction has not been possible for a period, for example because a system failure prevented data collection, the controller will *initialise* – setting all historical values of the CVs and MVs to the current value.

MPC supports the addition of disturbance variables (DV) or *feedforward variables*. These are variables which cannot be manipulated by the controller but affect the CV. They are included in the CV prediction and so effectively add feedforward control. Similarly any MV dropped is treated as a DV so that any changes made to it by the process operator are included in the prediction of the CV. Feedforward control can of course be added to PID type controllers but requires configuration of the DCS. MPC includes the feature as standard for all its CVs.

Basic controllers implement a fixed strategy. MPC permits more CVs than MVs and will select which CVs to control based on objective coefficients specified by the engineer. The MPC package will employ a linear program (LP) or similar algorithm to select the least costly (most profitable) strategy. Objective coefficients are applied to each MV and, in most packages, also to each CV.

PID controllers normally operate in dimensionless form with all inputs and outputs scaled as a fraction of instrument range. This is not the case for MPC; its process gain matrix is in engineering units consistent with the units of MVs and CVs.

The PID controller gain is entered as an absolute number with action defined as direct or reverse to take account of the sign of the process gain. Process gains used by MPC can be negative as required.

Although powerful, MPC packages should be applied intelligently. There can be a tendency to assume they can resolve almost any control problem. They can be costly. The first installation has to bear not only the licence cost but also the cost of the platform on which it is to run and engineer training. While the cost is often justifiable, consideration should first be given to less costly solutions – even if they do not capture all the benefits. To make this evaluation it is important to understand all the interactions before proceeding to the design stage.

Most MPC packages assume linear process behaviour. Much can be done outside the package to linearise variables by applying signal conditioning. MPC includes a feedforward function; however, this is a bias feedforward. If it makes engineering sense to apply ratio feedforward then MPC cannot achieve the same performance as a DCS-based ratio algorithm. If the feedforward DV is feed rate then, if it varies by more than ±20%, the use of ratio feedforward will ensure the process gains in the MPC matrix remain constant. The ratio could then become an MV of MPC if required.

At the current level of technology MPC packages do not scan frequently enough to handle very fast processes. For example it would be unwise to apply one to compressor anti-surge control unless there is a back-up fast scanning system to recover from surge.

The intelligent design would normally be a mixture of 'traditional' DCS-based techniques and MPC.

8.10 Parallel Coordinates

Parallel coordinates is two-dimensional graphical method for representing multiple dimensional space. In the example shown in Figure 8.26, a point in seven-dimensional space is represented by the coordinates (x_1, x_2, x_3, x_4, x_5, x_6, x_7). Since we cannot visualise space of more than three dimensions, the value of each coordinate is plotted on vertical parallel axes. The points are then joined by straight lines.

The technique is well suited to predicting the behaviour of a multivariable controller, even before step-testing has been started. Plant history databases comprise a number of instrument tag names with measurements collected at regular intervals. If we imagine the data arranged in a matrix so that each column corresponds to either a MV or a CV in the proposed controller and each row is a time stamped snapshot of the value of each parameter. To this we add a column in which we place the value of the proposed MPC objective function (C) derived from the values in the same row (where P are the objective coefficients for the m CVs, and Q the objective coefficients for the n MVs), i.e.

$$C = \sum_{i=1}^{m} P_i CV + \sum_{j=1}^{n} Q_j MV_j \tag{8.72}$$

Each row in the database is then plotted as a line on the parallel coordinates chart. The result will initially appear very confused with a large number of lines superimposed. The next step is to add the HI/LO constraints on each vertical axis. If a line violates any constraint on any axis then the whole line is deleted. The lines remaining will each represent an occasion in the past when all the process conditions satisfied all the constraints. The final step is to choose the line for which the value on the cost axis is the lowest. Since this axis is the MPC cost function, the line with the lowest value will represent the operation that

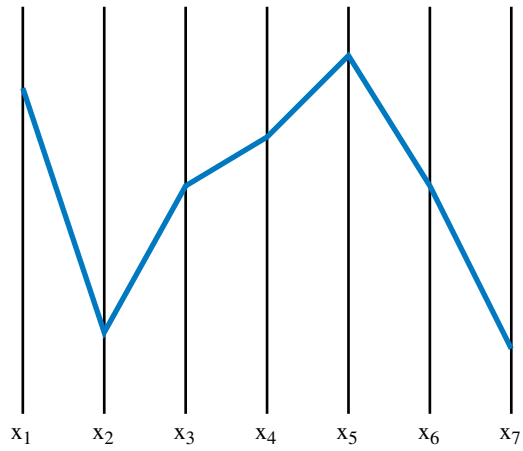

Figure 8.26 *Parallel coordinates*

MPC would select. Provided that the process has at some stage operated close to the optimum (as defined by MPC) then the chosen data set will give some idea of the operating strategy that MPC will implement. If different from the established operating strategy, this approach gives an early opportunity to explore why. Any difference should be seen as an opportunity to adopt a more profitable way of operating the process, rather than an error that should be corrected by adjusting the individual objective coefficients.

Another application is, for each variable, to plot a frequency distribution chart for all the operating data and superimpose on this a chart based on only the points lying within the chosen operating envelope. A significant narrowing of the range might indicate the need to improve basic regulatory control.

There are a number of commercially available packages that will allow data to be imported from the process history database and provide end-user tools to simplify its manipulation. Some include enhancements [26] such as automatically arranging the sequence of the axes, optimising the spacing and filtering out highly interdependent variables. These help support other applications of the technique, of which the main one in the process industry is the diagnosis of operating problems.

8.11 Enhanced Operator Displays

A major challenge with MPC is presenting its actions to both process operators and control engineers in a form that is readily understandable. Particularly with large controllers it is often difficult to diagnose why the controller is adopting a particular strategy. This can lead to the operator disabling the controller, or partly disabling it by tightening the MV constraints. Assuming there is no problem with the controller, such actions result in lost profit improvement opportunities.

While there is yet to be developed an entirely satisfactory solution to this problem, some ideas have been applied successfully. One is the use of a *radar plot*. This is similar to parallel coordinates except that the axes are arranged radially. Only a limited number of CVs and MVs are practicable – perhaps a maximum of around 12, so only the more important variables are included. Figure 8.27 shows a typical plot. The LO and HI limits are each plotted as a continuous polygon (shown as dashed lines), as is the current operation or predicted steady state (shown as the solid line). A specimen operation might also be included for reference. The human mind would appear to better at recognising shapes rather than tables of numbers.

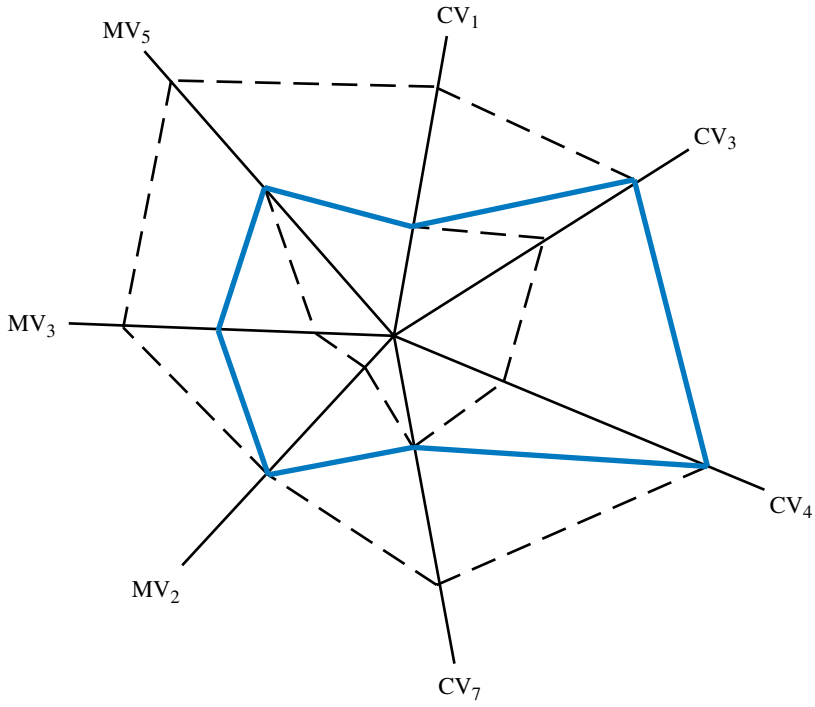

Figure 8.27 Radar plot

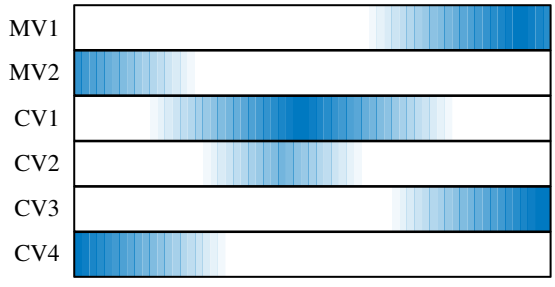

Figure 8.28 Heat map

With the use of colour to distinguish multiple plots, the change in shape is often readily recognised as normal or abnormal.

Another graphical approach is the *heat map* – shown as Figure 8.28. Each critical MV and CV is represented by a horizontal bar. Each bar is divided into small vertical slices. The extreme right hand slice of the bar represents the current situation. As a variable approaches its constraint its colour intensifies, for example between pale yellow and bright red, representing how the close the variable is to its constraint. At regular intervals the slices all move to the left, with the one on the extreme left being discarded. This diagram is helpful in showing how variables move in and out of constraint. Again the mind can recall patterns that are known to represent normal or abnormal behaviour.

8.12 MPC Performance Monitoring

The vendors of MPC packages offer increasingly sophisticated tools for monitoring the performance of their applications. The licence fees for such tools can be substantial and many of the functions included may not be seen as valuable by the engineer using them. The purpose of this section of the book is to present a number of ways in which performance might be monitored. Each technique has been applied somewhere, although not all at the same site. It is anticipated that the engineer will identify those that are valuable, decide whether to implement them and assess whether a proprietary package meets the needs.

This section describes a layered approach. At the top are simplistic overview tools primarily for management reporting. Below these the engineer can 'drill down' into increasing levels of detail to diagnose problems. If the engineer is to build the tools, such as in a spreadsheet package, they will need to retrieve information from the process data historian. In addition to the more conventional process measurements held in this database the tools will need the following:

- on/off status for each MPC
- the value of each MPC objective function
- the status of each MV and CV (i.e. have they been excluded or 'dropped'?);
- upper and lower limits set in the controller for each MV and CV, including hard and soft limits if applicable
- identification of the limiting constraints (not all MPC packages provide this so it may be necessary to build additional logic into the monitoring tools to flag those close to limits)
- the economic weighting factors (all MPC packages permit these on MVs, most also on CVs)
- either the unbiased prediction for each CV or the value of the bias used in the prediction

These requirements do increase the load on the historian but most of them compress well. Apart from the last item, the others change comparatively rarely. If the system supports data compression the incremental load will be very small.

Some sites find it important to maintain a high profile for advanced control to retain senior management's attention and commitment. In other sites the management demand regular reports. Performance can be condensed into a single number, i.e. the total benefits captured. However, it is important to remove from this number any changes outside the control of the site – for example changes in feed and product prices. A better approach is to report benefits captured as a fraction of the maximum achievable, as shown in Figure 8.29.

A portion of the available benefits may exist but there is insufficient return on investment to justify the cost of doing so. Rather than exclude them completely the chart includes a 'justifiable' value which is the maximum that could be captured. Future technological developments may bring down costs, and so what is not justifiable today should be occasionally reassessed rather than forgotten.

The 'installed' trend shows what could be captured if every installed control application is working 100% of the time. The gap between this and the justifiable value represents what new applications are required. The gap between the 'installed' trend and the 'achieved' trend represents what is being lost by applications not being fully utilised.

A quarterly management report including these trends supported by a few summary points can do much to facilitate improvements. Attention can be drawn to manpower shortages on the implementation team or application support. Critical unreliable instrumentation can be identified to support the case for replacement or prioritised maintenance.

While the chart deliberately excludes financial data, there is no reason why the recipient should not be made aware of to what 100% corresponds. An annual update of the true value of each application should

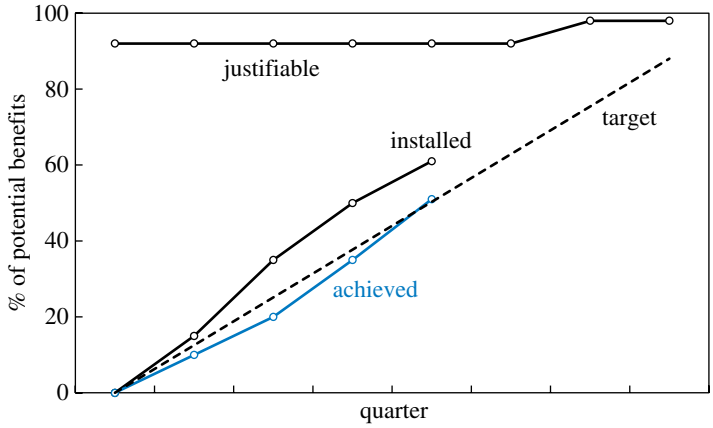

Figure 8.29 *Reporting benefits captured*

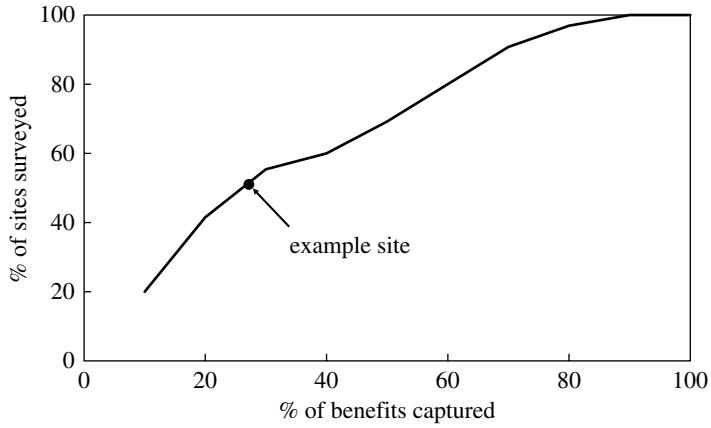

Figure 8.30 *Competitive positioning*

be completed and if this causes any change in any of the indices they can either be back-calculated for previous years or an explanation included in the report.

Reporting benefits as a fraction of what is achievable permits plant-to-plant or site-to-site comparisons. There are also consulting organisations that can provide comparison with competitors in a form such as that shown in Figure 8.30. The best site is capturing around 90% of the available benefits and the worst about 10%; the example site is capturing about 27%. In terms of its competition, of the sites surveyed, around 51% are doing better.

On sites with multiple production units, the next level of detail is a breakdown by unit. Figure 8.31 shows an example. For this the contribution that each unit makes is represented as a percentage of the total site benefits. Some form of service factor is required to quantify the uptime of each application. This might simply be the percentage of the time that the application is switched on. However, it is possible to constrain MPC so that it makes no changes and still be on 100% of the time. If required, a more complex, but more

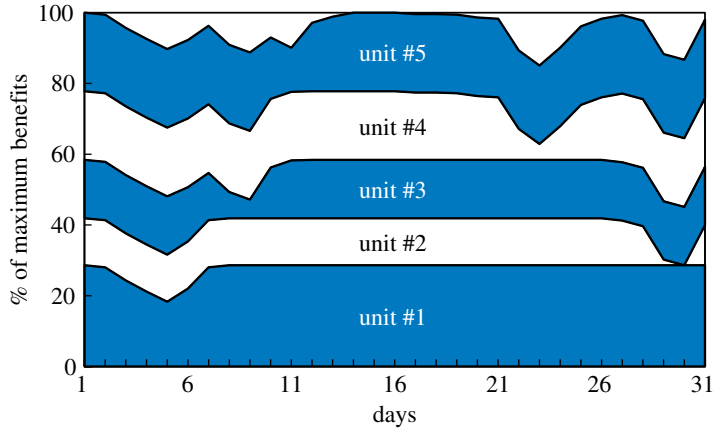

Figure 8.31 *Benefits breakdown by unit*

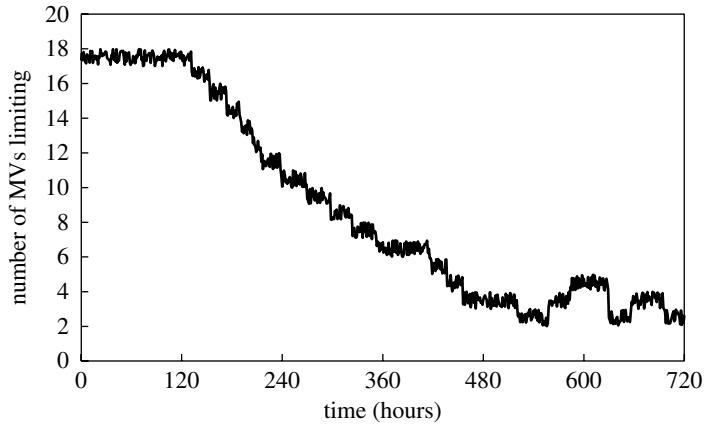

Figure 8.32 *Managing MV constraints*

realistic, definition of service factor can be used. This might be based on the proportion of MVs not against a constraint – possibly applying a weighting factor for each MV based on its economic importance.

The next level of detail is monitoring each MPC. It is possible simply to trend the controller's objective function. However, experience shows that this can be quite noisy and show discontinuities as constraints change. But the main concern with MPC is that it is over-constrained by the operator. It is common for the operators to periodically close the gap between the HI and LO limits on the MVs. This might be done temporarily for a good reason, because of some problem with the controller or the operator understanding of its actions, but the constraints are rarely relaxed again without some intervention by the engineer.

Figure 8.32 shows a trend over a period of about a month where efforts were made to remove as many MV constraints as possible. Some MPC packages generate engineer-accessible flags to identify whether a MV is limiting. If not, then the engineer has to develop some simple logic that checks whether each MV is close to a constraint. The flag is set to 1 if the MV is limiting and 0 if not. Any MV 'dropped' is treated as limiting. The flags are totalled and the result historised. At the beginning the controller was virtually disabled – able to manipulate only two MVs. This situation was reversed within the month.

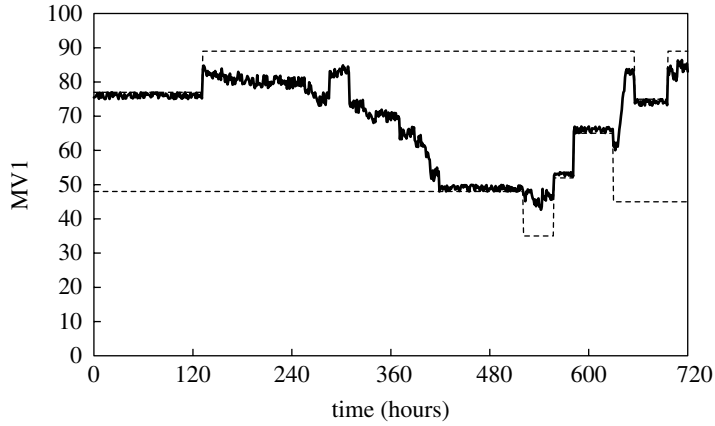

Figure 8.33 *Trend of a single MV*

Figure 8.33 shows one of the detailed trends used in support of the exercise. It shows, for MV1, the actual value and the HI and LO constraints. If the MPC package supports hard and soft constraints then both should be trended. The chart is useful in determining why the total number of constraining MVs has changed, and from the time of the constraint change, identifying who made the change and why.

Similar trending can be developed for CVs – although these tend to be less used. But detailed monitoring of the value of each CV is worthwhile. MPC internally calculates a predicted value for each CV from the MVs, using the dynamic models (G) obtained by step-testing.

$$CV_i = \sum_{j=1}^{m} G_{ij} MV_j + bias_i \tag{8.73}$$

Comparison is made between the predicted value and the actual value; the bias term is then adjusted to bring the two in line. The bias will always be nonzero since it is not true that the CV will be zero when all the MVs are zero. However, a large variation in the bias indicates a poor model. In order to monitor this it must be possible to retrieve either the bias term or *unbiased CV* from MPC. Figure 8.34 shows trends of data collected from MPC for both the unbiased and actual CV. To the eye the bias (the difference between the two values) appears constant. However, by trending the standard deviation of the bias we see in Figure 8.35 that this is not the case. Some event took place, approximately halfway through the collection period, to cause degradation in the accuracy of the CV prediction.

While this trend is an effective detection tool, it probably is not practical in this form. It is not immediately obvious whether the reduction in accuracy is sufficient to warrant attention. To address this we instead monitor the performance parameter (ϕ).

$$\phi = 1 - \frac{\sigma_{bias}^2}{\sigma_{CV}^2} \tag{8.74}$$

If the prediction is perfect then ϕ will have the value 1, since the bias will have remained constant and its standard deviation therefore zero. As the variation in the bias approaches the variation in the CV the prediction becomes increasingly valueless. To understand this, let us assume that the prediction of the

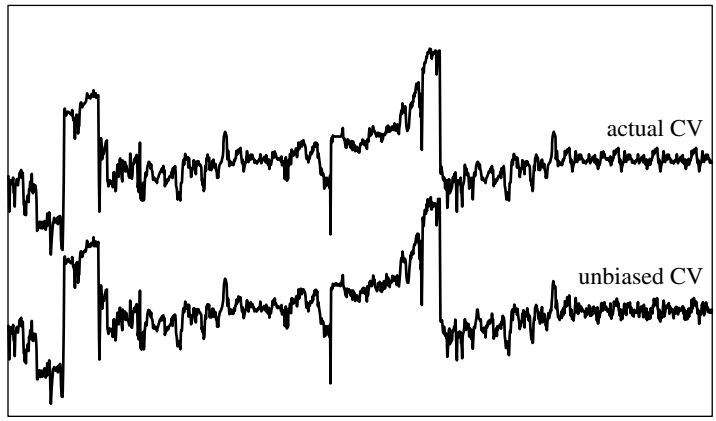

Figure 8.34 Assessing CV prediction error

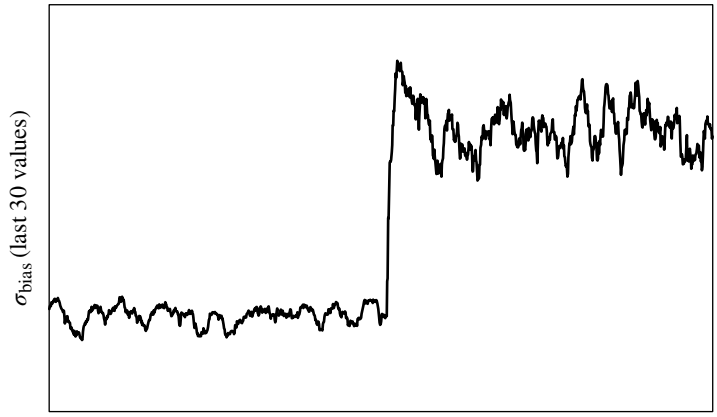

Figure 8.35 Trending the standard deviation of the bias

unbiased CV is that it is always constant. The standard deviation of the bias will therefore be the same as the standard deviation of the actual CV. The controller is effectively ignoring the prediction and ϕ will be zero. As ϕ falls below zero the prediction is so poor that it is creating disturbances greater than the natural disturbances in the CV. Using such a predicted CV is worse than assuming a constant CV and taking no corrective action. Most MPC projects are justified on the basis that the standard deviation in a limiting CV will be halved. For this to be achievable ϕ cannot be less than 0.75. In actuality, because some allowance needs to be made for other imperfections in the controller performance, it needs to be typically 0.9.

In its simple form ϕ is effective is assessing prediction accuracy prior to commissioning MPC but care must be taken in using it to monitor performance in real time. This is because its value may not change at the same time as a problem arising with prediction accuracy. If the prediction changes because of a change in process dynamics or a problem with one of the measurements on which it is based then there will be a significant change in σ_{bias} with perhaps little change in σ_{CV}. As a result ϕ will almost immediately reflect the problem. However, the change in bias can arise because of a change in CV due, for example, to an

unmeasured disturbance. In which case both σ_{bias} and σ_{CV} will increase – which may result in relatively little change in ϕ. To resolve this σ_{CV} can be calculated using data that is slightly older than that used for σ_{bias}. Consideration should also be given to the fact that successful MPC will reduce σ_{CV}, causing ϕ to decrease even though there is no degradation in accuracy. For example if ϕ is 0.75 before commissioning, and the controller successfully halves σ_{CV}, then its value will fall to zero. Both problems can be overcome by using a constant value for σ_{CV}, determined prior to commissioning of MPC.

Most MPC strategies have a large number of CVs and it would be unreasonable to expect the engineer to check all of the trends at frequent intervals. However, it is possible to generate an overall performance parameter, for example by trending the number of CVs which fail to meet the required performance. If this trend moves away from zero for significant periods then examination of the individual trends would identify the culprit(s).

Once a poor prediction has been detected we still have the problem of determining the cause. The prediction includes several dynamic models, any one of which may be the source of the inaccuracy. Further it could be caused by the absence of a model. By looking for correlations between ϕ and each of the MVs, the suspect model can be identified. This may simply be by eye – looking at trends of both. Or it may involve the use software, such as a spreadsheet package, to search for correlations between the prediction bias and each of the MVs.

If real process economics are used in MPC, then a wide range of other monitoring opportunities are created. These include:

- checking that MPC has truly optimised the process and not simply automated the existing operating strategy
- quantifying the lost opportunity if the operator over-constrains the MVs
- determining the value of debottlenecking projects
- calculating the benefit actually captured by MPC

Full details of how these techniques can be developed are included in Chapter 12 as a worked example on a simple distillation column.

9

Inferentials and Analysers

Accurate property measurement is key to the capture of many of the benefits of process control. Money can be made by more closely approaching product quality specifications. Process conditions can be continuously optimised, provided good product quality control is in place.

Property measurement falls into two basic categories:

- The first are mathematical techniques where basic process measurements of flow, temperature, level and pressure are used to infer a property. Often also called *soft sensors* or *virtual analysers*, they are used mainly to predict product quality but may be used for any parameter that cannot be measured directly – such as column flooding, catalyst activity, rate of coking, etc.
- The second is the use of on-stream analysers to directly measure product quality. It is not the intent of this book to cover any detail of how such analysers operate or how they should be installed or managed. Instead this chapter will focus on the use of their measurements in control strategies.

9.1 Inferential Properties

Even if a reliable on-stream analyser exists, it is usually still worthwhile to develop an inferential. Since the inferential is based primarily on basic measurements, it will respond much more quickly than the analyser. The analyser could well be located a long way downstream from the point at which the product is produced. Additional delay can be introduced by the sample system and the analytical technique employed. Figure 9.1 illustrates the benefit of this dynamic advantage on an example process. The two curves are each from an optimally tuned PID controller responding to the same process disturbance. Reducing the sample interval from 300 seconds, typical of an analyser, to 30 seconds results in a much smaller deviation from SP, sustained for a much shorter time.

Figure 9.2 shows the potential economic benefit. Point A represents a typical benchmark with a θ/τ ratio of 4. This might be from a process lag of 5 minutes and a deadtime of 20 minutes – both quite reasonable dynamics for a process such as a distillation column with a chromatograph on the distillate product rundown. In these circumstances an inferential could be expected to reduce the deadtime by at least 10 minutes (point B). Doing so would allow the controller to be tuned more quickly and would result in a reduction by about 33% in off-grade production.

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

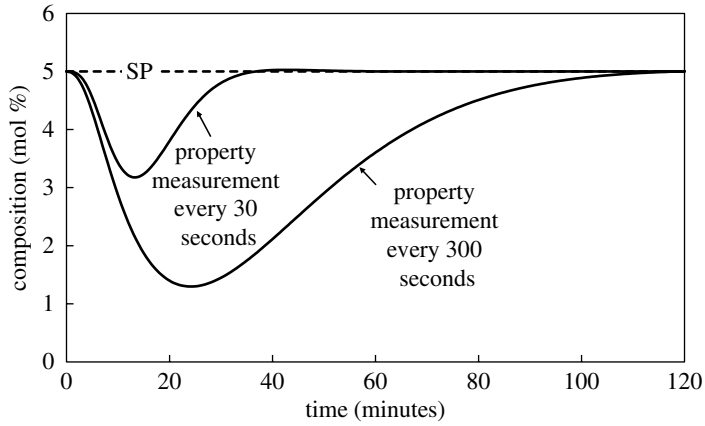

Figure 9.1 *Effect of early detection*

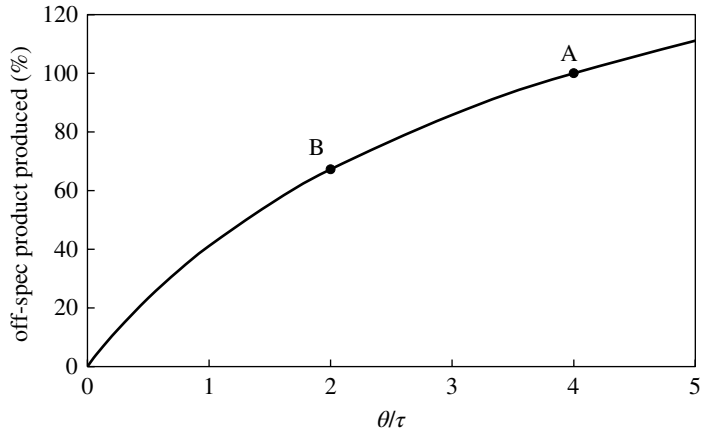

Figure 9.2 *Impact of deadtime on off-grade production*

Inferentials comprise a mathematical function (f) using a number of *independent variables (x)* to predict the value of a *dependent variable (y)*.

$$y = f\left(x_1, x_2 \ldots . x_n\right) \tag{9.1}$$

They fall into two groups – those derived from *regression analysis* of historical process data and *first-principle* types which rely on engineering calculations. First-principle techniques still require some historical data to calibrate the model and to check its accuracy. While the vendors of first-principle techniques might argue that the volume of data required is less, the key to the success of both techniques is the quality of the data. The use of routinely collected data, for example, from a plant history database, can often cause inaccuracies in the end result.

Firstly, the data may not have been collected at steady state. Not all the variables used in the inferential will have the same process dynamics. Shortly after a disturbance they will all be approaching steady state

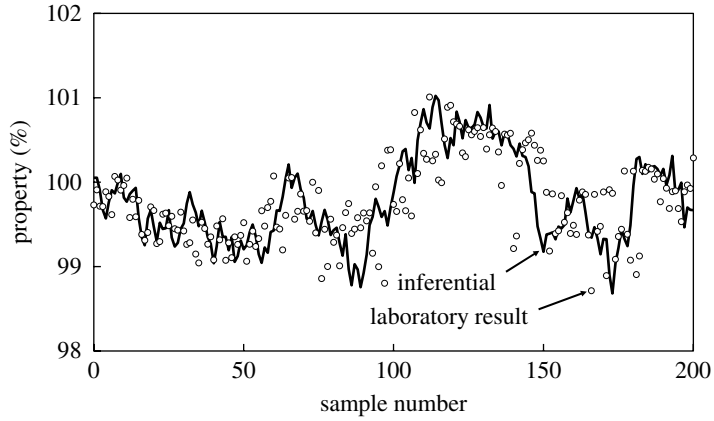

Figure 9.3 *Time-stamping error*

but to a different degree. The process may be temporarily out of energy or material balance as the inventory of either may be changing. Regression analysis is usually performed on data collected at a fixed interval, possibly averaging several sets of data around the collection time. While any errors introduced will be random and will not necessarily affect the form of the inferential, they will make it difficult to confirm its accuracy. With first-principle types, which may use only a few sets of data for calibration, it is more important that data are collected when the process is steady and has been steady for long enough for the deadtime of the dependent variable to expire.

Another potential problem is that of *time-stamping*. The dependent variable is often a laboratory result which may not be available until several hours after the sample was taken. It is therefore necessary to associate it with the operating conditions at the time of the sample. However, sample times are not necessarily reliable. Most sites will sample according to a schedule. However, the true sample time may be very different. It may have been delayed because there was an operating problem at the time or it may be taken early to fit in with the sampler's workload. Often all the samples on a process are scheduled for the same time but clearly could not all be taken simultaneously.

It is a misconception that, if the process is steady, recording the exact sample time is not important. Figure 9.3 is based on a hypothetical perfect inferential that exactly agrees with the laboratory result. The process is reasonably steady as seen by the trend of the inferential which varies less than ±1%. The mismatch of the laboratory samples is caused by introducing, into the sampling time, a random error in the range of ±10 minutes. Plotting the same information as a scatter chart, Figure 9.4, would suggest that a correlation which we know to be perfect is far from it. The error in prediction is comparable to the variation of the true value. If we were to develop an inferential from this information, we would have little confidence in its reliability. If we were monitoring the performance of an existing inferential, then we could be misled into disabling one that is working well.

While we could ask the sample taker to record the actual sample time, a more reliable approach is to automate this. One approach is to locate a push-button next to the sample point and connect it to the DCS so that it either logs the time when it is pressed or records all the independent variables at the time. Industries such as pharmaceutical manufacturing, where record keeping is of far greater importance, install sample points which record automatically the time that the sample valve is opened. It is also essential that the LIMS (laboratory information management system) has the facility to record actual rather than scheduled sample time.

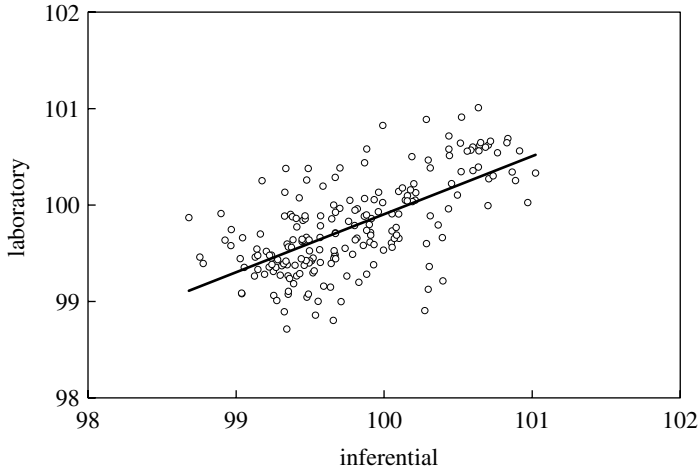

Figure 9.4 *Effect of time-stamping error*

Relying on routinely collected data will often not provide sufficient *scatter*. With modern data collection systems it is a trivial exercise to assemble information collected over several years. Even if a laboratory sample is only taken daily, assembling a thousand or more sets of data should present no problem. However unless the process is required to make multiple grades of the product, each with very different specifications, even without automatic quality control the process operator will have kept the quality very close to target. Any large deviations will usually be due to process upsets and may not provide any reliable steady-state information.

Figure 9.5 shows a typical situation where the development of an inferential, using only routine data, would be unreliable. Inclusion of only a few additional points, collected under test run conditions while moving in stages across a wide operating range, greatly improves accuracy. While regression analysis is generally thought to need 30 or more sets of data, 5 to 10 well-scattered, properly collected points will enable a reliable inferential to be developed. Confidence in the calibration of a first-principle model needs a similar amount of data.

It is common practice to collect such additional data during step-testing for a MPC project. Provided steady state is reached, then this will be useful to help validate an inferential. However, it may not be practical to cover all operating scenarios. For example, many inferentials are sensitive to feedstock but all types of feed may not be processed during the step-test phase. Secondly, it may prove impossible to develop an inferential from the data collected. It is too late a stage in the project to discover that additional instrumentation will be required.

First-principle models should be provided as *white boxes* with detailed documentation with perhaps *pseudo-code* that the engineer can convert to code appropriate to the DCS. This with its testing and documentation can be time-consuming. Inferentials can be *black boxes* that are delivered as compiled code with sparse documentation. They may require less implementation effort but can only be maintained by the supplier. Both types can be too complex to be properly understood by the engineer and so fall into disuse. Simple regression analysis tends to produce inferentials that are arithmetically simple and may therefore be readily built into the DCS using standard features. They can be described as *grey boxes*, particularly if they include some simple engineering principles. More complex regressed types, such as artificial neural networks, will require a separate platform and probably some proprietary software. They can therefore be more costly.

The decision on whether to use regression or first-principles technology is not straightforward – particularly if relying solely on the (less than impartial) vendors for information. We have already seen that both techniques require a similar amount of good quality data rather than a large volume of suspect data. Those supplying first-principle models will claim that regression analysis assumes that the input variables are independent and that true independence is unachievable. For example, distillation tray temperatures separated by a few trays will track each other closely. Indeed, as Figure 9.6 shows, even truly independent variables (such as distillation column reboiler duty and reflux flow) will show a correlation. In this case the column clearly generally operates with reboiler steam at around 14 t/hr and a reflux of around 57 m^3/hr. Significant increases in reboiler duty would cause operating problems unless reflux is also increased and so the variables appear not to be independent. While simple least squares regression will fail if one input variable is exactly correlated with another, experience shows that it will give good results even if there are cross-correlations. Where there are strongly correlated independents, regression will see little advantage

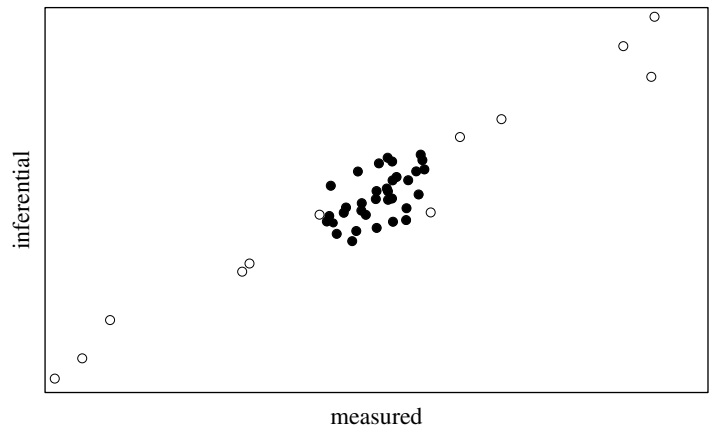

Figure 9.5 *Poorly scattered data enhanced with test run results*

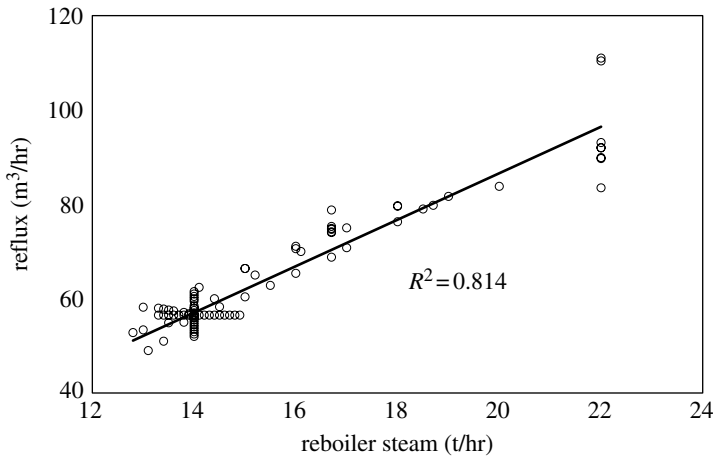

Figure 9.6 *Correlation between apparently independent variables*

in using both. If there is some advantage to using one over the other, e.g. because it is a more reliable measurement, then it would be wise to manually exclude the other from the analysis. A good statistical analysis package will identify such cross-correlations and indicate the improvement in accuracy of the inferential that is achieved as each variable is added. There are also many other techniques which do not assume independence. Further, the so-called first-principle techniques can include correlations developed by others by regressing experimental data.

Models based on engineering principles should theoretically adapt more readily to minor process modifications. This would mean that they could be used, unlike regression, without waiting for additional process data to be collected. However, they are rarely 'pure' and often include calibration factors. It would be a brave engineer that trusted them implicitly without re-checking the calibration.

Regression analysis is open to abuse if applied without an understanding of the process. For example, blindly applying an artificial neural network effectively discards any knowledge of process behaviour. While the resulting inferential may work well, its performance outside the range over which it was trained can be extremely unpredictable. There are examples where this has caused a reversal of the sign of the process gain with respect to the key MV – severely impacting process profitability.

Naïvely applying linear regression techniques can have a similar impact. With modern spreadsheets and statistical packages, it is relatively easy to extract large quantities of data from the process information database and search for all possible correlations. By including a large number of process variables and a wide range of arithmetical transformations (such as powers, logarithms, ratios, cross-products, etc.), it will certainly be possible to apparently improve the accuracy of the inferential. However, this is likely to be only a mathematical coincidence. Consider the example where we attempt to predict a value (y) from one input (x_1) such that

$$y = a_0 + a_1 x_1 \tag{9.2}$$

If we have three sets of data (or records) from which to derive this correlation, then a_0 and a_1 would be chosen to give the best fit to the available data. Imagine that this does not give the required accuracy and so we introduce a second input (x_2). Since we have only three records, then the coefficients (a_0, a_1 and a_2) could be determined by solving the following equations simultaneously.

$$(y)_1 = a_0 + a_1 (x_1)_1 + a_2 (x_2)_1 \tag{9.3}$$

$$(y)_2 = a_0 + a_1 (x_1)_2 + a_2 (x_2)_2 \tag{9.4}$$

$$(y)_3 = a_0 + a_1 (x_1)_3 + a_2 (x_2)_3 \tag{9.5}$$

This would then give us a perfect fit. However, this 'perfection' would be achieved even if x_2 was a random number completely unrelated to y. Of course, we generally have far more data sets than independent variables but this illustrates the point that the inclusion of any additional input will appear to improve the correlation. The relationship between the degree of correlation and the number of inputs is approximately linear. So, for example, if the number of inputs is half the number of records we would achieve 50% of perfection even if the inputs are random numbers. Statistical techniques used to determine the true value of each input are included in Chapter 14. Effective inferentials typically have between one and three inputs, although some of these may be of the *compound* type where an input is derived from two or more measurements.

If, during the development of an inferential, the inclusion of such a compound variable is found to contribute greatly to accuracy, then consideration should be given to modifying the basic control scheme

to use this value. For example, an inferential for the composition of a distillation product will often include a tray temperature (T) and column pressure (P), taking the form

$$Q = a_0 + a_1 T + a_2 P \tag{9.6}$$

This can be rewritten as:

$$Q = a_0' + a_1 T' \tag{9.7}$$

We can compensate the temperature for deviations from the mean pressure ($\bar{P}$):

$$a_0' = a_0 + \frac{a_2}{a_1}\bar{P} \quad \text{and} \quad T' = T + \frac{a_2}{a_1}\left(P - \bar{P}\right) \tag{9.8}$$

More detail is given in Chapter 12, but T' is the basic form of a *pressure compensated temperature (PCT)*. Instead of incorporating pressure in the inferential, T' could be calculated in the control system and used as the PV of the tray temperature controller. This would more quickly respond to changes in column pressure. It may also render the inferential unnecessary and so resolve another issue. If the measurements used in an inferential are also measurements used by basic controllers then, instead of the PV, we can use the SP as the input to the inferential. This will be noise-free and also stops the inferential taking unnecessary corrective action during a disturbance that will soon be resolved by the basic controller. PV tracking will result in the inferential effectively using the PV if the controller is switched to manual. The disadvantage of this approach is that the inferential-based composition controller will then manipulate the tray temperature SP but this, because this SP will also be a key input to the inferential, will result in an instantaneous change in its PV. We have effectively cascaded one controller to another with slower dynamics – resulting in tuning difficulties. Implementation of the PCT approach can avoid the need for the cascade.

Another example of a compound variable might be an inferential predicting conversion in a reactor based on a number of catalyst bed temperatures.

$$Q = a_0 + a_1 T_1 + a_2 T_2 + a_3 T_3 \tag{9.9}$$

One of these temperatures is likely to be the PV of a basic controller. It can be advantageous to replace this PV with the *weighted average bed temperature (WABT)* which has the form

$$WABT = x_1 T_1 + x_2 T_2 + x_3 T_3 \quad \text{where} \quad x_1 + x_2 + x_3 = 1 \tag{9.10}$$

The x coefficients can be calculated from:

$$x_i = \frac{a_i}{a_1 + a_2 + a_3} \tag{9.11}$$

Assuming T_3 is the reactor outlet temperature, we might make this the temperature controller using a PV that is compensated for T_1 and T_2 deviating from their means ($\bar{T}_1$ and $\bar{T}_2$). This would be given by:

$$T_3' = T_3 + \frac{a_1}{a_3}\left(T_1 - \bar{T}_1\right) + \frac{a_2}{a_3}\left(T_2 - \bar{T}_2\right) \tag{9.12}$$

Care may be necessary if the residence time in the reactor is large so that the bed temperatures have very different dynamics. Dynamic compensation could then be applied to T_1 and T_2 – effectively incorporating bias feedforward from the upstream temperatures, as described in Chapter 5.

If the inferential includes terms which make no engineering sense (or coefficients which have the wrong sign), then it will fail during a process excursion. However, there is also the risk of excluding terms that appear at first not to make engineering sense. Chapter 12 gives some examples where nonlinear transformations, ratios and cross-products can make sense, as can coefficients with apparently the wrong sign.

If the inferential is to be a CV of MPC, then care needs to be taken with applying regression analysis to derive a linear function. Consider the MPC gain matrix shown as Equation (9.13).

$$\begin{pmatrix} K_{11} & K_{12} & . & K_{1n} \\ K_{21} & K_{22} & . & K_{2n} \\ . & . & . & . \\ K_{m1} & K_{m2} & . & K_{mn} \end{pmatrix} \begin{pmatrix} MV_1 \\ MV_2 \\ . \\ MV_n \end{pmatrix} = \begin{pmatrix} CV_1 \\ CV_2 \\ . \\ CV_m \end{pmatrix} \tag{9.13}$$

MPC will thus predict CV_1

$$CV_1 = K_{11}MV_1 + K_{12}MV_2 + \ldots + K_{1n}MV_n + bias_1 \tag{9.14}$$

A linear inferential will have the form

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n \tag{9.15}$$

If y is used as CV_1 and x_1 is MV_1, x_2 is MV_2, and so on, then it is important that a_1 is equal to K_{11}, a_2 is equal to K_{12}, etc. Since the inferential's coefficients are derived from regression and the process gains subsequently derived from step-testing, they are unlikely to be exactly the same. If there are other inputs to the inferential that are not included in the gain matrix, then the need for exact agreement will depend on whether those inputs change if an MV is changed.

If the inferential uses process measurements that are physically far apart, then a process disturbance may affect one measurement more quickly than another. As a result the inferential may show complex dynamic behaviour, such as inverse response. If used as the measurement of a PID controller, then the slow tuning necessary to maintain stability may give very poor control. While, in theory it is possible to dynamically compensate the inputs, the compensation required will depend on the source of the disturbance. One source may cause input 1 to change before input 2, while another may cause the reverse. If the inferential is to be a CV of MPC, then such packages can handle high order dynamics such as inverse response. However, they too will be prone to the dynamics changing depending on the source of the disturbance. In a regression type inferential it is straightforward to exclude the less critical input if its dynamics are very different and repeat the regression analysis without it. Some accuracy will be sacrificed, but controllability will be greatly improved. In a first-principle model the simplest solution is to assume a constant value for the offending measurement.

The pragmatic approach to selecting the technology is to choose the approach that works better in each situation. Regression analysis usually has the lower cost and can be performed by the plant owner using a spreadsheet package or a proprietary development tool. If regression fails to deliver an inferential of sufficient accuracy, then a first-principle approach can be explored. This is likely to require a specialist supplier that, if truly convinced of the technology they offer, should be prepared to work on 'no win, no

fee' basis. If their product cannot outperform the benchmark established by regression, then they would waive their charges. This leads us to the question of how we assess the accuracy of an inferential.

9.2 Assessing Accuracy

While there a number of good statistical methods for assessing accuracy, they need to be applied with care. We address here, and later in Chapter 14, techniques applicable specifically to steady state data. This is the commonest form of validation, often comparing the inferential with the laboratory result.

It is very common for engineers to use the wrong statistic and so draw a false conclusion concerning the inferential's suitability for use in a controller. However, before going into detail of such statistics, we should keep in mind that the whole point of an inferential is to give a dynamic advantage. A delay of several hours waiting for the result of a laboratory test, or even several minutes for an on-stream analyser, limits how well we can control the property. It is often the case that the main disturbance to product properties is a change in feed composition. We need to ask whether an inaccurate inferential is better than no inferential. An inferential that can predict the effect of feed composition might offer a considerable benefit, provided its prediction is in the right direction and that any error in the prediction is less than the actual change in property. Provided some method is included that later updates the inferential, for example, a laboratory sample taken at steady state, then its inaccuracy could be acceptable. Before rejecting an inferential because it has not met some accuracy criterion for its steady state performance, we should consider whether the dynamic advantage is sufficiently beneficial that it outweighs other problems.

Figure 9.7 shows the method, favoured by suppliers of inferentials, for demonstrating the performance of an inferential. Line charts tend to lead one to believe the correlation between the inferential and the actual property is better than it is. Presenting the same data in Figure 9.8 as a scatter plot gives a more precise measure. For example, if the true property is 50%, the inferential will be between 30 and 70% – possibly far too inaccurate to be of any value.

The other favoured approach is the use of the statistic known as Pearson R^2 (described in more detail in Chapter 14). If there are n sets of data where x and y are the two variables that might be correlated, this is defined as

$$R^2 = \frac{\left(\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i}^{n}(y_i - \bar{y})^2} \tag{9.16}$$

If y is the measured property and $\hat{y}$ the predicted value, then an alternative definition is

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i}^{n}(y_i - \bar{y})^2} \tag{9.17}$$

This makes it clearer that R^2 is the proportion of the total variance in the measured property that is explained by the variance in the prediction. It is however, inadvisable to use this equation for time-dependent data. For example, if the sequence of the values of $\hat{y}$ was randomly shuffled (leaving the sequence of y

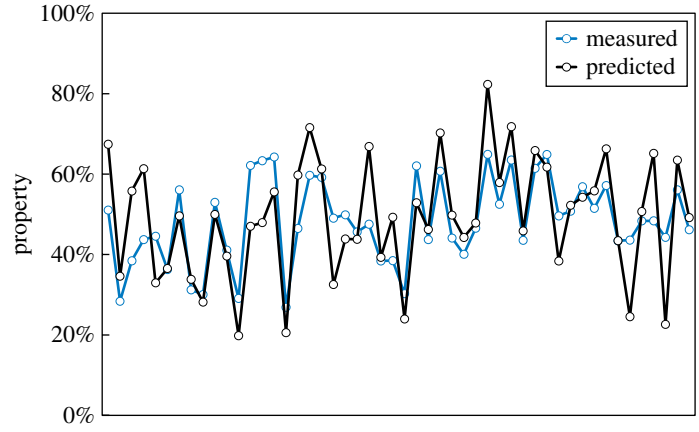

Figure 9.7 *Use of line plot to validate inferential*

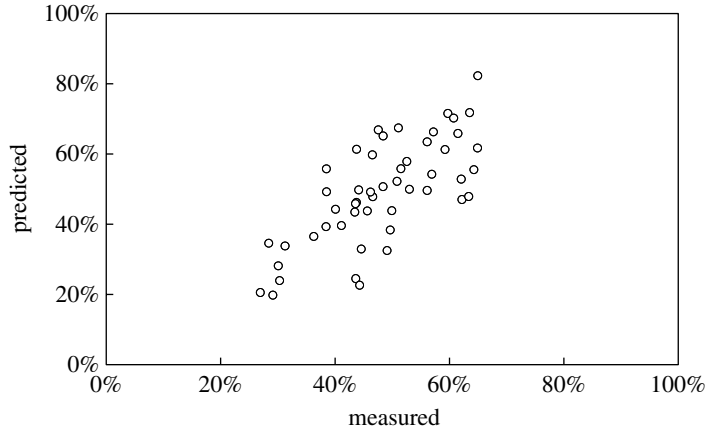

Figure 9.8 *Use of scatter plot to validate inferential*

unchanged), this would not affect the value derived for R^2 despite there no longer being a correlation between the two parameters. This problem does not arise with Equation (9.16).

A perfect correlation would have a value of 1 for R^2. However, a value close to 1 does not necessarily indicate that an inferential is useful. As an illustration, consider the graph shown in Figure 9.9 for the stock price of a process control vendor. Figure 9.10 shows the performance of an inferential developed by the author. With R^2 of 0.99 one would question why the developer is not a multi-billionaire. The reason is that it failed to predict the large falls in the value of the stock. The three occasions circled undermine completely the usefulness of the prediction. The same is true of an inferential property. If there is no change in the property then, no matter how accurate, the inferential has no value. If it then fails to respond to any significant change then it may as well be abandoned.

A better approach is to compare the standard deviation of the prediction error (σ_{error}) against the variation in the actual property (σ_{actual}). We show in Chapter 13 that benefit calculations are usually based on the assumption that the standard deviation of the actual property is halved. If we assume that our control

Figure 9.9 *Process control vendor stock price*

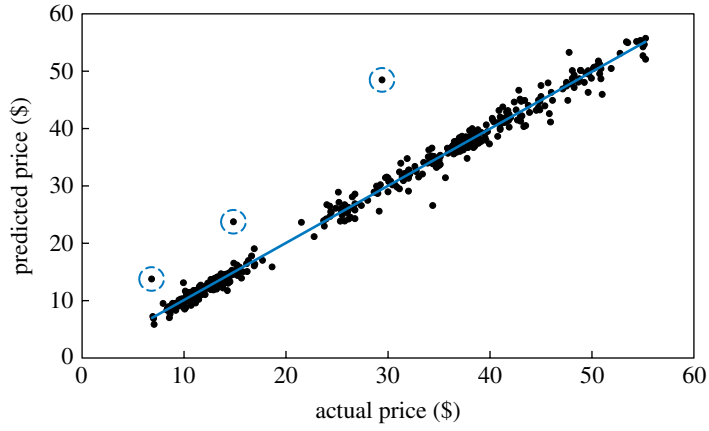

Figure 9.10 *Predicting stock price*

scheme is perfect and the only disturbance comes from the random error in the prediction then, to capture the benefits

$$\sigma_{error} \leq 0.5\,\sigma_{actual} \tag{9.18}$$

This can be written in the form of a performance parameter *(ϕ)*:

$$\phi = \left(1 - \frac{\sigma^2_{error}}{\sigma^2_{actual}}\right) \geq 0.75 \tag{9.19}$$

We first introduced this parameter in Chapter 8, using it to assess whether the CV predictions made by MPC were sufficiently reliable. It clearly has a value of 1 when the inferential is perfect. To understand

how it works over its full range, consider the inferential $Q = a_0$ where a_0 is the mean of all the property measurements used to build the inferential. The inferential will have a bias error of zero. But, since the inferential always generates the same value, the standard deviation of the prediction error will be the same as the standard deviation of the actual property. The inferential clearly has no value and the value of ϕ will be zero.

Next consider the case when the true property does not change. Any error in the prediction will cause the controller to wrongly take corrective action. In general, if the standard deviation of the prediction error is greater than that of the true property, ϕ will be negative – indicating that the inferential is so bad that process performance would be improved by switching off the controller. Figure 9.11 trends this parameter for the stock price example. It confirms what we know, that the prediction will lose us money on several occasions.

A further limitation of the use of R^2 is that, if there is a perfect relationship between inferential ($PV_{inferential}$) and laboratory result ($PV_{laboratory}$), the value of R^2 will also be 1 for any linear function, i.e.

$$PV_{inferential} = a_1 PV_{laboratory} + a_0 \tag{9.20}$$

So, for example, if a_1 had a value of 3 and a_0 a value of 0, then the inferential would be treble the laboratory result but, according to the statistical test, be working perfectly. The same would apply if a_1 were negative – even though this reverses the sign of the process gain.

To illustrate the difference between R^2 and ϕ, consider the data in Table 9.1. Column 1 is a series of laboratory results. Columns 2 to 4 are the corresponding results from three inferentials derived using different values of a_1 and a_0. The values of ϕ in columns 2 and 3 of the table confirm, unlike R^2, that the inferential would be so poor that its use would cause control of the property to worsen. In column 4, where only a bias error is introduced, both R^2 and ϕ show that the inferential would be perfect – requiring just a once-off correction for it to be useful.

The parameter (ϕ) can be used both in the development of an inferential and (with modification) its monitoring. At the development stage we clearly need its value to be greater than 0.75 but, given that this assumes perfect control, in reality it needs to be higher if we are to capture the benefits claimed. A more realistic target is 0.9.

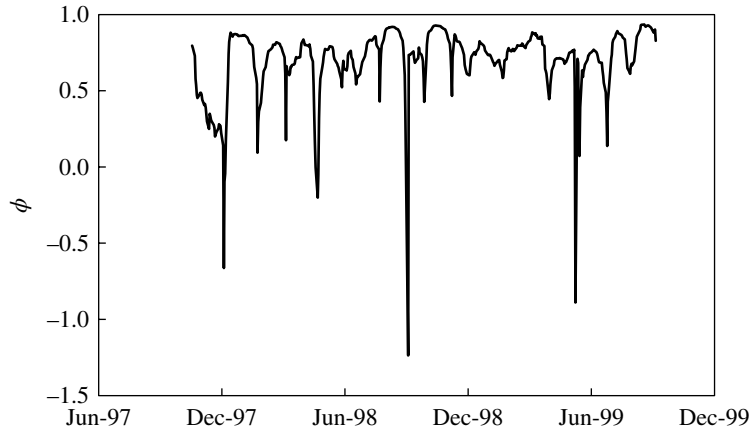

Figure 9.11 *Inferential performance parameter*

Table 9.1 Comparison between R^2 and ϕ

$PV_{laboratory}$	$a_1 = 3$ $a_0 = 0$	$a_1 = -1$ $a_0 = 0$	$a_1 = 1$ $a_0 = 5$
4.81	14.43	–4.81	9.81
4.79	14.37	–4.79	9.79
5.25	15.75	–5.25	10.25
5.02	15.06	–5.02	10.02
4.86	14.58	–4.86	9.86
4.96	14.88	–4.96	9.96
5.08	15.24	–5.08	10.08
5.17	15.51	–5.17	10.17
4.98	14.94	–4.98	9.98
4.90	14.70	–4.90	9.90
4.86	14.58	–4.86	9.86
4.98	14.94	–4.98	9.98
4.94	14.82	–4.94	9.94
5.17	15.51	–5.17	10.17
5.01	15.03	–5.01	10.01
5.17	15.51	–5.17	10.17
5.09	15.27	–5.09	10.09
5.16	15.48	–5.16	10.16
4.75	14.25	–4.75	9.75
4.81	14.43	–4.81	9.81
ϕ	–3	–3	1

It is common to attempt to improve the accuracy of an existing inferential by collecting more recent process data. However, if the existing inferential has been at least partially successful in improving control, this will have caused a reduction in σ_{actual} and result in a lower value of ϕ. If the history is available, σ_{actual} would be better calculated from data collected before the original inferential was commissioned. This would also permit ϕ to be determined for the existing inferential and therefore any improvement quantified. It would also permit the economic impact of the new inferential to be assessed. If ϕ_1 and ϕ_2 are the before and after values, then the potential percentage increase in benefits captured is given by

$$100\left(1-\sqrt{\frac{1-\phi_2}{1-\phi_1}}\right) \tag{9.21}$$

This formula might also be used to justify enhancing (or replacing) a poorly performing inferential with an on-stream analyser.

If ϕ is calculated at a high frequency, e.g. by the use of on-stream analyser measurements, then care must be taken to ensure that the process is at steady state. Because the dynamics of the analyser will be longer than those of the inferential, any change in the inferential will be reflected some time later in the analyser measurement. There will therefore appear to be a transient error, even if both the inferential and analyser are accurate. Alternatively, dynamic compensation can be applied. We cover this later in this chapter.

We might wish to take account of known variation caused by the laboratory testing. Most laboratory tests follow a documented standard, which will include estimates of *repeatability* (r) and *reproducibility*. Reproducibility is not of concern here. It relates to the agreement between results obtained from different

laboratory instruments, different technicians and different laboratories. Repeatability however, is of interest. It relates to agreement from the same technician re-testing the same sample using the same laboratory instrument. This is defined, for example, by the ASTM, as

$$r = 2\sqrt{2}\sigma_{laboratory} \tag{9.22}$$

This gives us the variance of the error in the laboratory test.

$$\sigma^2_{laboratory} = \frac{r^2}{8} \tag{9.23}$$

Since an error in the laboratory test affects both the actual value used in the calculation of ϕ and also the prediction error, the calculation would be modified to

$$\phi = 1 - \frac{8\sigma^2_{error} - r^2}{8\sigma^2_{actual} - r^2} \tag{9.24}$$

However, we are only predicting laboratory error based on published standards. We do not know the true error. The laboratory will often perform better than the quoted repeatability but there are also many other sources of random error; so the accuracy of the result could also be worse than the quoted repeatability. Further, if the true process variation is then less than the assumed laboratory repeatability, the value of ϕ will exceed 1. Any statistical parameter we choose for monitoring purposes will be subject to these problems. A pragmatic approach is to define ϕ according to Equation (9.19), determine how effective it is and only then explore the impact of testing errors as necessary.

As a monitoring tool ϕ can be very valuable in the early detection of degradation in the accuracy of an inferential and disabling it before its poor performance does any real harm. However, it needs to be used with care.

- If our controller is successful, it will reduce σ_{actual}. Our performance parameter will then fall, misleadingly indicating that the performance of the inferential has degraded. For example if, at the design stage, σ_{error} was half of σ_{actual}, then ϕ would have a value of 0.75. If the controller successfully halves σ_{actual}, then ϕ will drop to zero – suggesting the inferential is of no value. To avoid this we choose a constant value for σ_{actual}, equal to the variation before the controller was commissioned.
- The observant reader may have noticed that the sudden drops in ϕ in Figure 9.11 do not occur at the same time as the unforeseen drops in stock price. So while the technique may be effectively used to assess accuracy at the design stage, it has little value in this form as a monitoring tool. The problem arises because, on the day that the stock priced drops, there is a large increase in not only the variance of the error but also in the variance of the actual value. Their ratio therefore changes little. This too would be resolved by using a constant value for σ_{actual}. The large error is not therefore associated with a large change in the actual value and the value of ϕ will show a spike at the same time as the error occurs.
- We have to use a number of historical values to calculate σ_{actual} – usually 30. Thus, even if a problem with the inferential is resolved, the performance index will indicate a problem until 30 more laboratory results are taken. While we can reduce the number of historical values used, a better approach would be to treat as outliers the occasion(s) where the inferential is now known to have failed and remove them from the calculation of the index.
- Finally we should recognise that a failure may not be due to a problem with the inferential but a problem with the laboratory result. This leads us on to our next topic.

9.3 Laboratory Update of Inferential

With well-integrated information systems it is relatively easy to automatically update the inferential with the latest laboratory result. Any difference between the laboratory result and the value of the inferential at sample time can be used to update the bias term in the inferential calculation. Because of the delay in obtaining the laboratory result, we do not know the current error (E_n) but we can determine the previous value as

$$E_{n-1} = \left(PV_{inferential} - PV_{laboratory}\right)_{n-1} \tag{9.25}$$

The bias may be then updated as

$$bias_n = bias_{n-1} - KE_{n-1} \tag{9.26}$$

K is a filter parameter set by the engineer to a value between 0 and 1. There is a natural reluctance to set it to 1 since this would accept the full correction immediately; it may be the laboratory result that is in error. By setting it to a lower value, typically around 0.3, several results will be required for the full correction to be made. If the error remains constant, then the bias ultimately required is given by

$$bias_{required} = bias_{n-1} - E_{n-1} \tag{9.27}$$

Using Equation (5.29) to apply a first order exponential filter to the bias update

$$bias_n = P.bias_{n-1} + \left(1-P\right)\left(bias_{n-1} - E_{n-1}\right)$$

$$\text{or} \quad bias_n = bias_{n-1} - \left(1-P\right)E_{n-1} \tag{9.28}$$

Comparing Equations (9.26) and (9.28) shows that P is the same as 1–K. Equation (5.30) shows therefore that the update passes through a first order lag of $-ts/\ln(1-K)$, where ts is the laboratory sampling interval. Assuming K has been set at 0.3 and the laboratory samples are daily, then the lag will be about 67 hours.

In general the number of samples (n) required for the required fraction (f) of the error to be eliminated is given by

$$n = \frac{\log\left(1-f\right)}{\log\left(1-K\right)} \tag{9.29}$$

Rearranging gives

$$K = 1 - \left(1-f\right)^{1/n} \tag{9.30}$$

Figure 9.12 shows the effect the choice of K has on the number of samples required to eliminate 90% of the error. For example, with K set to 0.32 it would take six samples.

It is possible to optimise the value for K. Using historical data the update can be built into a spreadsheet and K adjusted to minimise the sum of the squares of the error. However, in most cases the optimum value of K will be found to be zero. While updating in this way would seem a good idea, in almost every case it causes the accuracy of the inferential to degrade. The laboratory result is subject to error. To this must be

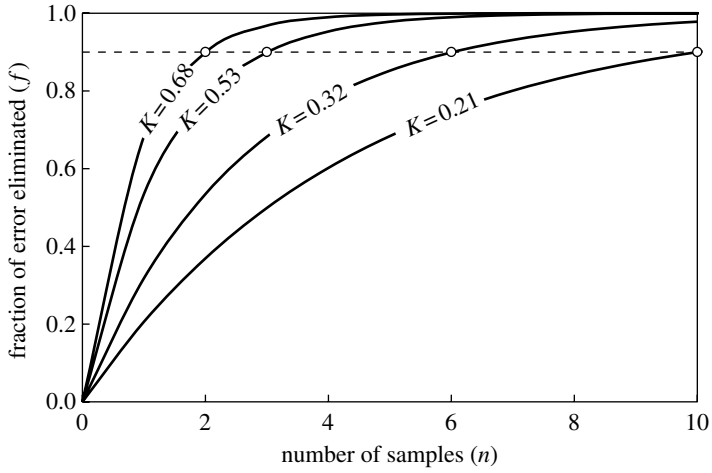

Figure 9.12 *Effect of K on elimination of bias error*

added many other sources of error such as time-stamping, sample contamination and human error. The variance of the laboratory result is already included in the variance of the inferential error – since this is defined as the difference between the inferential and the laboratory. Passing this variance also through the bias update increases the variance of the inferential error by the factor $(1 + K^2)$. Hence the standard deviation (σ_{error}) will increase by the square root of this factor.

The problem is that we need to distinguish between *bias error* and *random error*. What we have described so far are random errors. A bias error is a systematic difference between the true value and its measurement. It is unlikely to exist in the laboratory result but can arise in the inferential. A change of feedstock may cause a bias error to arise. In the oil refining industry, for example, it is common to have *cold property* specifications on fuels – such freeze point, cloud point and pour point. These are controlled by changing operating conditions on the process but are also affected by the *paraffinicity* of the crude oil from which the product is derived. Changing the type of crude being processed can therefore cause a bias error in the inferential. In the chemical industry it is common to infer quality of a product based on operating conditions in the reactor in which it is produced. However, as the catalyst activity declines over time, a bias error will accumulate in the inferential.

The best solution to a bias error is to eliminate it at source. If we can achieve this, then we can abandon completely any automatic updating of the inferential. In our example it may be possible to detect the change of feedstock or possibly rely on an operator to enter the change in the DCS. There are techniques for compensating for changes in catalyst activity, e.g. by including in the inferential a parameter representing the total volume of feed processed – maybe weighted by a measure of severity.

If we do need to update automatically, then we need to separate the bias error from the total error. The *CUSUM technique* offers an effective solution. In this case CUSUM is the cumulative sum of differences between the inferential and the laboratory result. Table 9.2 presents an example calculation.

Provided the results are in the correct sequence, there is no need for the sample interval to be fixed. Thus if samples are taken at irregular intervals, such as repeat tests, they may still be included. Figure 9.13 shows the CUSUM trend. If the error were 100% random, the trend would be noisy but horizontal and no bias update is required. If a bias error is present, then the slope of the CUSUM trend is the amount by which the inferential is overestimating and so the amount by which the bias should be reduced. We showed

Table 9.2 *CUSUM calculation*

Sample	Inferential	Laboratory	Error	CUSUM
1	5.08	4.81	0.27	0.27
2	4.97	4.79	0.18	0.45
3	4.93	5.25	–0.32	0.13
4	5.05	5.02	0.03	0.16
5	5.20	4.86	0.34	0.50
6	5.55	4.96	0.59	1.09
7	5.22	5.08	0.14	1.23
8	5.52	5.17	0.35	1.58
9	5.56	4.98	0.58	2.16
10	5.56	4.90	0.67	2.82
11	5.64	4.86	0.78	3.61
12	4.80	4.98	–0.18	3.43
13	5.16	4.94	0.23	3.65
15	4.95	5.17	–0.22	3.43
16	4.93	5.01	–0.09	3.35
17	4.95	5.17	–0.22	3.13
18	5.17	5.09	0.08	3.21
19	5.17	5.16	0.01	3.22
20	5.16	4.75	0.41	3.63

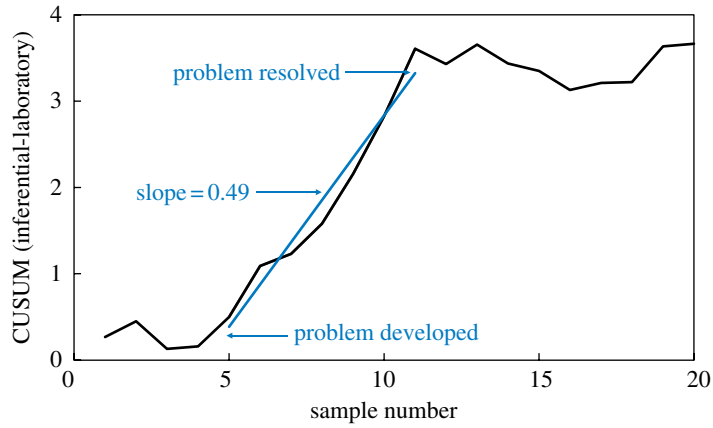

Figure 9.13 *CUSUM trend*

in Chapter 5, when developing the least squares filter, that the slope can be determined from the last N values of the CUSUM (where Σ_N is the most recent and Σ_1 the oldest)

$$\text{slope} = B_1\Sigma_N + B_2\Sigma_{N-1} \ldots . + B_r\Sigma_{N-r+1} \ldots . + B_N\Sigma_1 \tag{9.31}$$

where the coefficients (B) are determined using Equation (5.69) – repeated here:

$$B_r = \frac{6\left(N-2r+1\right)}{N\left(N^2-1\right)} \tag{9.32}$$

In our example, choosing a value 5 for N gives a slope of 0.49. Since it already includes several historical values the correction can applied immediately with confidence. It might be argued that waiting for the CUSUM to develop a noticeable trend delays correction, in this example, by five samples. However, we saw from applying Equation (9.29), with K set at 0.32, conventional updating would take six samples to implement only 90% of the correction, while additionally amplifying any random error.

It is important to record that a bias correction has been made so that subsequent estimates of the CUSUM slope do not include values collected before the correction. This can be achieved either by retrospectively applying the correction to the results collected since the problem developed or by resetting the CUSUM to zero.

It is possible to automate calculation of the slope of the CUSUM line and use this to automatically update the bias. We would first need to decide how many historical values (N) are used in the calculation. Setting N to 2 in Equation (9.32) gives a slope of $(\Sigma_N - \Sigma_{N-1})$ but this is simply E_n (the most recent prediction error) and so is equivalent to setting K to 1 in Equation (9.26).

Increasing N to 3 gives a slope of $(\Sigma_N - \Sigma_{N-2})/2$. This, perhaps surprisingly, omits Σ_{N-1}. In fact the slope of the CUSUM line is unaffected by the middle of the last values. This arises from the assumption that laboratory samples are collected at a fixed interval and occurs for any odd value for N. Choosing $N = 3$ results in the mean of the last two errors ($n = 2$) being used to update the bias.

$$bias_{new} = bias_{old} - \frac{E_n + E_{n-1}}{2} \tag{9.33}$$

Increasing n to 3 gives

$$bias_{new} = bias_{old} - \frac{3E_n + 4E_{n-1} + 3E_{n-2}}{10} \tag{9.34}$$

In general, for any choice of n

$$bias_{new} = bias_{old} - \frac{6}{n(n+1)(n+2)} \sum_{r=1}^{N-1} r(n-r+1)E_{n-r+1} \tag{9.35}$$

The bias update is therefore a weighted average of the last n errors. It would be possible, using the data that were used to build the inferential, to optimise the choice of n. Increasing n will cause less amplification of random error but will delay bias update. However, rather than update the bias after every sample, a better approach is to monitor the slope of the CUSUM and only update if this exceeds some chosen value. Further, rather than calculating the weighting factors, they too should be optimised. Removing the constraint that they should sum to 1 would also allow the best choice of K to be included. Indeed retaining K as 1 will result in instability as n is increased.

If the error has even a small random component, the performance index (ϕ) will always worsen if automatic updating is implemented. The index measures only random error. If there was no random error, then the index would have a value of 1 – no matter how large the bias error. The advantage of the CUSUM approach is that it reduces the impact of the random error and so the effect that updating has on ϕ is much reduced.

Whether the correction should be fully automated is debatable. It is certainly important to monitor random and bias errors frequently but automatic correction is not a substitute for a poor inferential. Its inclusion may disguise a problem. Indeed, this is exactly the situation with our predicted stock values. The prediction was quite simply yesterday's value with automatic updating based on today's value.

Care needs to be taken to ensure that updating the inferential does not 'bump' the process. A sudden change in measurement may cause too rapid a change to the manipulated variable(s). This can be overcome by forcing the controller to reinitialise or by ramping in the correction. Care also needs to be taken if the process operator, after seeing the laboratory result, has already responded to any error in the inferential by adjusting the target. If so, this could be dealt with by changing the target in step with the change in bias.

If the inferential was developed by regression, then common practice is simply to add bias updating without redesigning the inferential. Strictly the coefficients used in the inferential should be regressed with bias updating in place. To understand why this is the case, consider what would have happened if the bias error was somehow corrected at source by, for example, modifying the values collected for the dependent variable. If these modified values were then used to build a new inferential, it would include quite different coefficients. Indeed, it is possible for the sign of a coefficient to change.

9.4 Analyser Update of Inferential

Automatic updating using an on-stream analyser measurement is quite different from updating with laboratory results. Analysers can have a reputation of poor reliability but we describe later in the chapter techniques that prevent spurious measurements from disturbing the process or being used to update an inferential. With this measurement validation in place analysers are far less prone to random errors than the laboratory. Secondly analysers provide measurements far more frequently and so the delay introduced by filtering will be far less. Referring again to Figure 9.12, an analyser, such as a chromatograph measuring the content of a single component, might update every five minutes. Choosing 0.32 for K would result in 90% of any discrepancy between inferential and analyser being eliminated within 30 minutes.

However, we have already mentioned the problem that the process dynamics introduce. We could resolve this by only permitting updates when the process is at steady state and has been so long enough for the analyser to respond to any changes on the process. However, processes rarely reach a true steady state and updates are likely to be fairly infrequent. Instead we can install the configuration shown in Figure 9.14.

We apply dynamic compensation in the form of a deadtime/lead-lag algorithm. This is tuned in exactly the same way as described in Chapter 6 covering bias feedforward. By performing open loop steps on the MV we obtain the dynamics of both the inferential and the on-stream analyser. Applying Equation (6.23) we get

$$K = -\frac{\left(K_p\right)_{analyser}}{\left(K_p\right)_{inferential}} \tag{9.36}$$

The process gain of the analyser and the inferential should be the same and so K should be –1. If the test shows that this is not the case the problem should be resolved before commissioning analyser updating – indeed before using the inferential in a controller.

From Equation (6.25)

$$\theta = \theta_{analyser} - \theta_{inferential} \qquad T1 = \tau_{inferential} \qquad T2 = \tau_{analyser} \tag{9.37}$$

The analyser deadtime should be significantly larger than that of the inferential; otherwise the inferential serves little purpose – except perhaps as a back-up in the event of analyser failure. So θ will be positive. If not the case, the dynamic compensation should be applied to the analyser measurement.

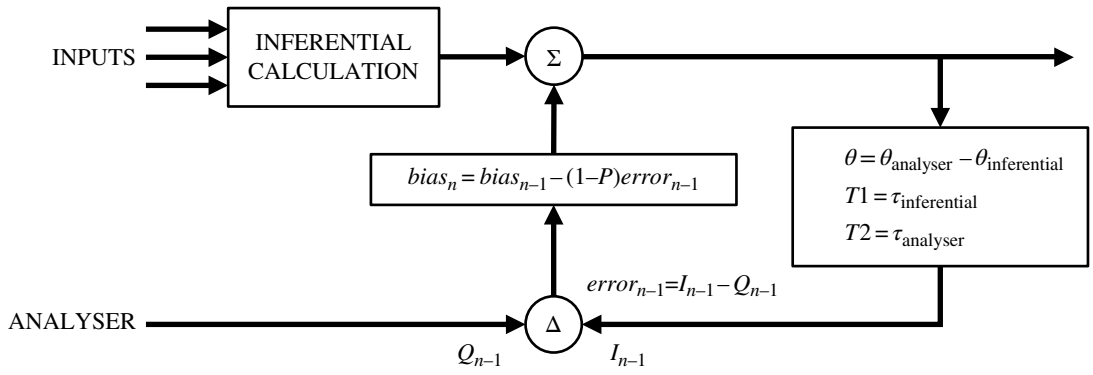

Figure 9.14 *Use of analyser to update inferential*

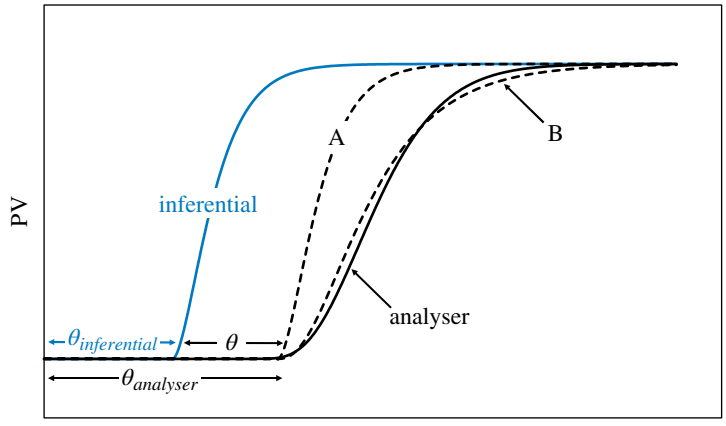

Figure 9.15 *Effect of dynamic compensation*

If the analyser is discontinuous and its sample interval greater than the time it takes the process to reach steady state, then it may not show significant lag. As $T2$ should not be set to zero (because of the effect on the $T1/T2$ ratio), then it is wise only to include the deadtime compensation – by removing the lead-lag or setting $T1$ equal to $T2$.

The way in which the dynamic compensation operates is shown in Figure 9.15. The inferential is shifted (to curve A) by the delay θ. This compensates for the difference in deadtime between the analyser and the inferential. The lead term $(T1)$ cancels out the lag in the inferential and the lag term $(T2)$ replaces it with the lag of the analyser, changing the output (to curve B). This now closely matches the analyser.

The correction term is the difference between the dynamically compensated inferential and the analyser measurement. The dynamic compensation assumes first order behaviour and so is unlikely to be exact. Further there will be inaccuracies in estimating the values of the time constants. This will cause an apparent error in the inferential but, providing it has the same process gain as the analyser, will be transient. Rather than correct for the error instantly a small exponential filter (a lag) is included in the bias update. If the analyser is discontinuous then, between measurements, an error will exist. Again this is transient

and will disappear at the next measurement. A substantially heavier filter will be required (with P set to around 0.98). Or, to avoid this, updating could be better configured to occur only when the analyser generates a new value – in which case P can be reduced to around 0.7.

The bias used by the inferential should be monitored. Since the updating forces the inferential and the analyser to always agree at steady state, a problem with either measurement will be not be obvious. An increase in the standard deviation of the bias will indicate a problem caused by random error.

9.5 Monitoring On-Stream Analysers

Many of the monitoring techniques suggested for inferentials can be applied to analysers. For example, the performance index (ϕ) can be used to identify excessive random error between analyser and laboratory. The CUSUM can be used to check for a bias error which can arise particularly if the analysis method does not exactly match the laboratory technique.

However, before applying such techniques, we should first try to minimise the sources of error. For example, locating the laboratory check sample point close to the analyser will minimise the time difference between taking the analyser reading and taking the check sample. For discontinuous analysers with a long sample interval an external indicator showing that a new sample is being taken can be used by the sampler so that the check sample is taken at the same time.

If in addition to calibration samples, routine samples are taken of the same product then these should ideally be taken from the same point. If the previously suggested push-button or automatic detection is installed then the analyser can also be checked against accurately time-stamped laboratory samples.

Analyser sample delay should be minimised by locating the analyser as close as possible to the process and installing a *fast loop.* This takes a small stream from a high pressure point in the process, routes it close to the analyser and returns it to the process at a point where the pressure is lower. A common approach is to connect the fast loop between the discharge and suction of a product pump. It is not advisable to install the loop around a variable pressure drop, such as a control valve, since the sample deadtime will then vary and cause controller tuning problems. If necessary, a fast loop pump can be installed. The analyser sample is taken from the fast loop. The sample should be taken as far upstream as possible, again to reduce delay. Vapour travels faster than liquid so taking a sample while still in the vapour phase, or vaporising it at source, will further reduce delay – but the sample lines will then need to be heated and insulated. Otherwise the heavier components will condense before the sample reaches the analyser and so affect the result.

The choice of analyser technology may be a trade-off between accuracy and speed of response. We will see in Chapter 10 that fuel gas heating value can be approximately inferred from its specific gravity or derived accurately from a full chromatographic analysis. SG analysers can be installed to give almost no delay, while chromatographs will delay the measurement by many minutes.

Analyser *sample conditioning* should be designed to ensure the sample is ‘clean’ and in the same condition as that when processed by a laboratory instrument. These recommendations and many others are covered by specialists [27].

There will inevitably be a difference between analyser and laboratory. Organisations have adopted a variety of approaches to resolving this. Placing responsibility for the accuracy of both devices under the laboratory manager prevents long debate about which result is correct. Moving towards the exclusive use of analysers for product certification raises their profile and the level of management attention given to their maintenance.

Close monitoring permits poorly performing analysers to be identified and the evidence provided to justify their improvement or replacement. It also provides evidence to dubious process operators that a

previously suspect analyser is now reliable. In addition to such historical monitoring, it is important to check the performance in real time of an analyser being used closed loop. A single undetected failure can result in costs greater than the annual benefit of improved quality control. Process operators and the plant manager will remember, for a long time, the incident of a whole batch of off-grade product that had to be reprocessed or downgraded. This can damage the reputation of all analysers; it takes far more effort to establish a good reputation than it does to destroy one.

PV validation is a technique which can be applied to any measurement but is of particular importance to analysers. A number of checks can be made and automatic control or inferential updating disabled if any of these fail. Firstly, the analyser may itself generate alarms. As an addition, sensibly set high/low checks on the measurement will flag a measurement that has moved outside its normal range. This is usually a standard feature within the DCS. If there is an inferential we can use the maximum expected deviation from the analyser to continuously update the high/low checks. The DCS might also offer rate-of-change checking. A measurement moving faster than the process dynamics permit would also be declared invalid. We need also to check for a low rate of change or 'frozen' value. This can occur with failure of discontinuous analysers employing *sample-and-hold*. While a low rate of change check would detect this, it is also likely to generate spurious alarms if the process is particularly steady. A better approach is a *timeout* check. Most discontinuous analysers provide a *read-now* contact that can be connected as digital input to the DCS. This is used to initiate a countdown timer to a value slightly higher than the sample interval. If this timer reaches zero the analyser is assumed to have failed.

By configuring a tag for each analyser, set equal to 1 if the measurement is valid and to 0 when not, we can historise this tag and use it to trend analyser availability and to average it as required. We can also set up similar tags to monitor the time that each analyser is in automatic control. This information then forms the basis of analyser performance reporting.

We also have to consider what action is taken on restoration of a valid measurement. If the analyser has been out of service for some time then the best approach is to generate a message to the process operator that it can now be restored to automatic control. If the outage is brief, then automatic re-commissioning might be considered, ensuring that correct initialisation is triggered to ensure the process is not 'bumped' by the measurement being different from that last used by the controller.

10

Combustion Control

This chapter confines itself to boilers and other fired process heaters that burn liquid or gaseous fuels, or a mixture of both. Fuel gas in particular can be major source of process disturbances – particularly if its pressure or composition can vary. In the case of mixed firing it may not be possible to manipulate the flow of all the fuels, e.g. because one may be a by-product, from another part of the process, that cannot be economically stored.

Fired heaters will normally have at least a feedback controller maintaining the required outlet temperature. Boiler duty will often be manipulated to maintain the required steam header pressure. However, if there is more than one boiler supplying the header, it is common to operate one or more as *baseload* boilers, where the duty is fixed and the remainder as *swing* boilers – used to control the steam pressure.

While the control strategies in this chapter are largely presented for heaters (and heater outlet temperature control), the majority are also applicable to boilers (and steam pressure control). Where there are differences, these are described.

10.1 Fuel Gas Flow Correction

Assuming gas flow is measured using a conventional orifice plate type of flow meter; we covered in Chapter 5 the correction necessary if working in units of normal volumetric flow, i.e. measured at standard conditions.

$$F_{true} = F_{measured}\sqrt{\frac{MW_{cal}}{MW}\times\frac{P}{P_{cal}}\times\frac{T_{cal}}{T}} \tag{10.1}$$

Or, if working in mass flow units

$$F_{true} = F_{measured}\sqrt{\frac{MW}{MW_{cal}}\times\frac{P}{P_{cal}}\times\frac{T_{cal}}{T}} \tag{10.2}$$

However, we also mentioned that special attention is required when applying these formulae to fuels. This is because a change in molecular weight not only affects meter calibration but also the heating value of the

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

gas. If Equation (10.1) is used to condition the measurement of a flow controller, then an increase in the molecular weight will cause the PV to fall and the flow controller to compensate by opening the control valve. Heating value generally increases with molecular weight and so, to maintain a constant fired duty, we need the control valve to close. The addition of compensation has worsened the impact of the disturbance caused by the change in fuel composition.

It also presents a potential safety hazard. It is common to ratio combustion air flow to the fuel flow measurement. Thus the increasing fuel gas MW would cause a reduction in air flow. As Figure 10.1 shows, using some common fuels as examples, this is opposite to what is required. There is thus the danger of combustion becoming sub-stoichiometric. The resulting loss of combustion efficiency would cause the outlet temperature (or steam pressure) to fall and the controller to increase fuel further.

Before incorporating heating value into the controller we need to ensure we use the correct definition. *Gross heating value (GHV)* is the heat released per unit of fuel if any water, produced by combustion, is condensed and so releases its heat of vaporisation. *Net heating value (NHV)* is a lower value because it is based on the water remaining as vapour. Both can be quoted on a volumetric or on a weight basis. The combustion products of most fired heaters leave as flue gas and so we will use NHV. As an energy-saving measure condensing heaters are likely to become more common. The only effect on the control schemes covered in this chapter will be a change of coefficient.

On-stream analysers measuring molecular weight are normally marketed as *densitometers* and so we will base the control design on *specific gravity (SG)*, defined as

$$SG = \frac{MW}{MW_{air}} \tag{10.3}$$

The molecular weight of dry air (MW_{air}) is generally assumed to be 28.96, as derived from Table 10.1. Figure 10.2 shows the relationship between NHV (on a weight basis) and SG for a number of gases commonly found in fuel gas.

Provided the fuel gas comprises mainly hydrocarbons, then its heating value (on a weight basis) is largely independent of composition. Thus calibrating the gas flow meter to measure mass flow and correcting using Equation (10.2) will result in negligible disturbance to the process as the gas composition changes.

However, if hydrogen is present in a significant proportion, then this approach will fail because of its very different heating value. Gas compositions are generally quoted on a volume (or molar) basis.

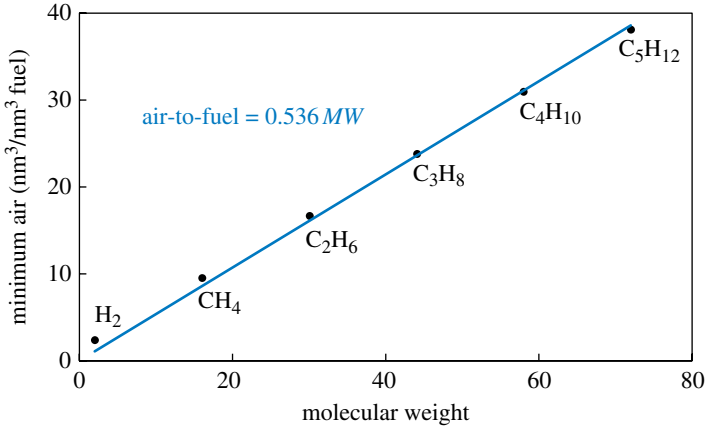

Figure 10.1 *Combustion air requirement*

Table 10.1 *Molecular weight of air*

Gas	Vol %	MW
O_2	20.95	32.00
N_2	78.09	28.01
CO_2	0.03	44.01
Ar	0.93	39.94
air	100.00	28.96

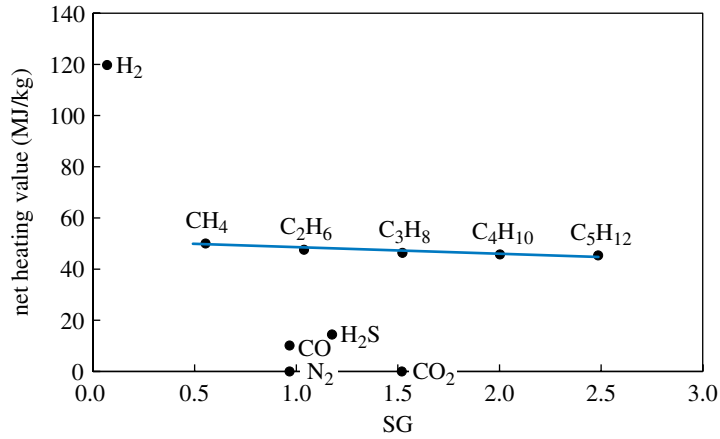

Figure 10.2 *Net heating value on a weight basis*

Figure 10.3 shows the relationship between weight % and volume % for hydrogen mixed with gases commonly found in fuel. It shows that (on a volume basis) the hydrogen content needs to be significant before moving away from simple mass flow control of fuel gas. This is confirmed by Figure 10.4 which shows that low levels of hydrogen have little impact on heating value. At levels above around 35 vol% we have to adopt a different approach.

Figure 10.5 shows the relationship between NHV (now on a normal volume basis) and SG. On this basis the heating value of hydrocarbons now varies with SG but in a way which can be inferred from

$$NHV = aSG + b \tag{10.4}$$

The coefficients can be derived theoretically based on the NHV and SG of pure gases. In the engineering units used for the graph, a is 56.14 and b is 5.78. If NHV is measured in units of BTU/SCF, then these coefficients change to 1506 and 139 respectively. However, these values should be used only as a guide. The true value will depend on what other components are in the fuel. The presence of inerts such as N_2 and CO_2 will change the relationship, as will any non-hydrocarbon fuels such as CO and H_2S. Provided the concentration of these components is small or varies little we can still predict NHV from SG but we need to develop the correlation from real process data. Figure 10.6 shows some typical laboratory data routinely collected from a site's fuel gas system over several months.

At first glance the correlation would appear to be poor with one point (ringed) showing a very large deviation. However, a review of the analysis of this sample, shown in Table 10.2, reveals a common problem

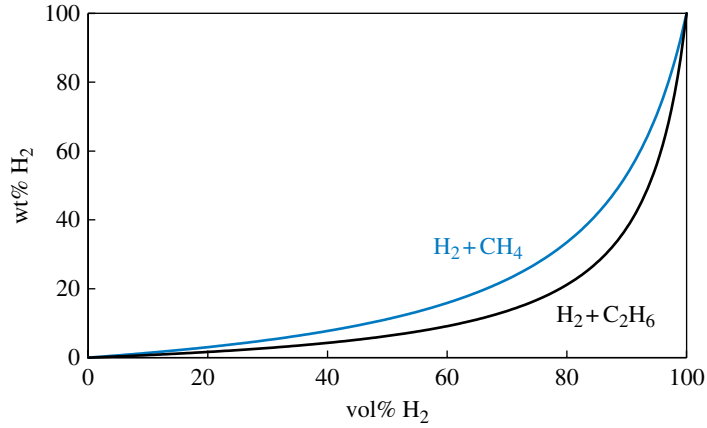

Figure 10.3 *Conversion of volume % to weight % hydrogen*

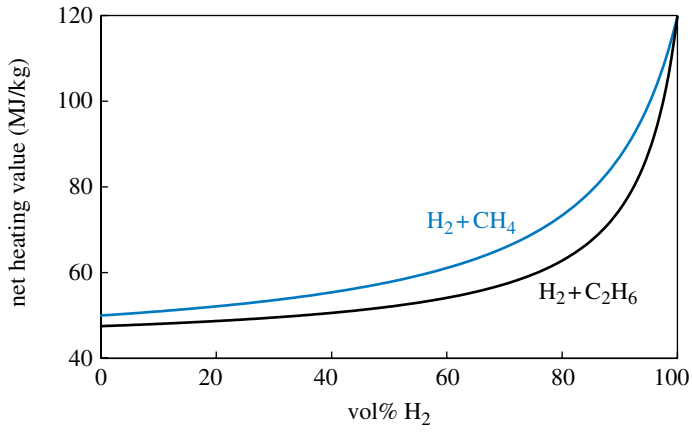

Figure 10.4 *Impact of hydrogen on heating value*

with sampling. Poor procedures have resulted in the sample being contaminated with air, as indicated by the O_2 content of 2.70%. We can remove this component from the analysis provided we also remove the associated N_2. We know from Table 10.1 that the N_2 concentration in air is 3.73 times that of O_2 and so we reduce the N_2 by 10.06%. The remaining 1.84% N_2 is that genuinely in the fuel gas. Applying this correction to every sample gives the very reliable correlation shown in Figure 10.6.

Once we have values for a and b we can modify the signal conditioning so that the flow is measured in energy units, e.g. MJ/hr or BTU/hr. By combining Equations (10.1), (10.3) and (10.4) we get

$$F_{energy} = F_{measured}\sqrt{\frac{SG_{cal}}{SG}\times\frac{P}{P_{cal}}\times\frac{T_{cal}}{T}}\times\left(aSG+b\right) \tag{10.5}$$

We can see that, if the molecular weight (represented now by SG) increases, the PV of the flow controller will now increase. The controller will respond now by closing the control valve as required.

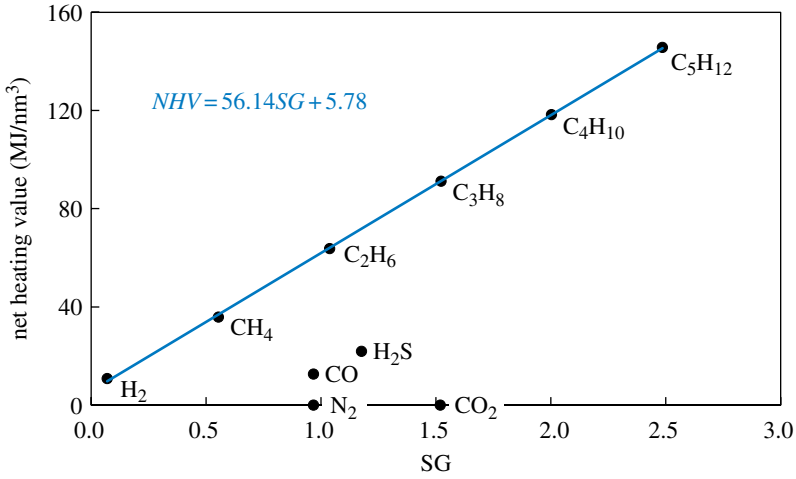

Figure 10.5 *Gas heating value on normal volumetric basis*

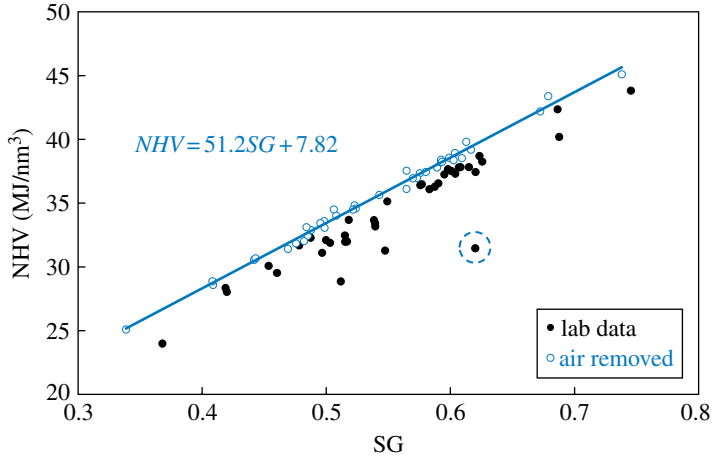

Figure 10.6 *Predicting NHV from plant data*

For meters measuring the actual volumetric flow, as described in Chapter 5, the calculation of duty can be derived from Equations (5.26) and (10.4).

$$F_{energy} = F_{actual}\sqrt{\frac{P}{P_{standard}} \times \frac{T_{standard}}{T}} \times \left(aSG + b\right) \tag{10.6}$$

For meters, such as the coriolis type, that directly measure the mass flow of the gas we combine Equations (5.28), (10.3) and (10.4).

$$F_{energy} = F_{mass} \times \frac{R.T_{standard}}{P_{standard}.MW_{air}} \times \frac{aSG + b}{SG} \tag{10.7}$$

Table 10.2 *Correcting fuel gas analysis*

Gas	Original analysis				Corrected analysis			
	MW	SG	NHV	mol %	MW	SG	NHV	mol %
H_2	2.02	0.07	10.8	46.90	2.02	0.07	10.8	46.90
CH_4	16.04	0.55	35.8	12.10	16.04	0.55	35.8	12.10
C_2H_6	30.07	1.04	63.7	10.10	30.07	1.04	63.7	10.10
C_2H_4	28.05	0.97	59.0	1.70	28.05	0.97	59.0	1.70
C_3H_8	44.10	1.52	91.2	6.70	44.10	1.52	91.2	6.70
C_3H_6	42.08	1.45	85.9	1.20	42.08	1.45	85.9	1.20
C_4H_{10}	58.12	2.01	118.4	3.80	58.12	2.01	118.4	3.80
C_4H_8	56.11	1.94	113.0	0.70	56.11	1.94	113.0	0.70
C_5H_{12}	72.15	2.49	145.3	1.10	72.15	2.49	145.3	1.10
C_6H_{14}	86.18	2.97	172.0	0.30	86.18	2.97	172.0	0.30
O_2	32.00	1.10	0.0	2.70	32.00	1.10	0.0	0.00
N_2	28.01	0.97	0.0	11.90	28.01	0.97	0.0	1.84
CO	28.01	0.97	12.6	0.50	28.01	0.97	12.6	0.50
CO_2	44.01	1.52	0.0	0.20	44.01	1.52	0.0	0.20
H_2S	34.08	1.18	21.9	0.03	34.08	1.18	21.9	0.03
total	17.96	0.62	31.5	99.93	16.37	0.57	36.1	87.16

Some plant owners prefer not to replace the conventional flow measurement with one recording in energy units. They argue that the process operator should be able to see the measurement in its original units. This can be displayed separately or, instead of conditioning the PV, we can apply the reciprocal of the function to the SP.

Figure 10.7 trends the SG used to derive the coefficients *a* and *b*. The daily variation can be as much as 0.2. If correction were not applied, to an orifice type flow meter, this would result in a change of fired duty of around 20%. Of course, from daily laboratory results, we do not know how quickly the SG changes. Even the largest change, if ramped smoothly over 24 hours, would easily be handled by the feedback controllers on the site's heaters and boilers. However, it is likely that changes will be more rapid and, in any case, there are often other reasons why we want to measure fired duty rather than simple fuel flow.

- In this example, the change in NHV would cause the process gain, between heater outlet temperature (or boiler steam production) and measured fuel flow, to vary by a factor of ±20%. The tuning of the feedback controller will therefore be away from optimum for long periods. This also applies to other control schemes, described later in this chapter, including heater inlet temperature feedforward and flue gas oxygen control.
- Any change in fired duty requires a change in combustion air. It is common therefore to install an air-to-fuel ratio scheme. While the required air-to-duty ratio remains constant, as fuel composition changes, the required air-to-fuel ratio does not. It is quite likely that a large undetected increase in NHV will result in sub-stoichiometric combustion and cause a major process incident.
- On heaters and boilers operating close to process limitations, such variation in fired duty could temporarily cause violation of a constraint – at least until the feedback controller returns the duty to that required. On baseload boilers there is usually no feedback controller to bring the boiler back within its operating limit.

Figure 10.8 shows the percentage error in the NHV, predicted from the measurement of SG, compared to the true value. As anticipated, from Figure 10.6, the prediction is quite accurate. The largest error is around 2% – probably acceptable even if there is a rapid change in fuel gas composition. It certainly would have

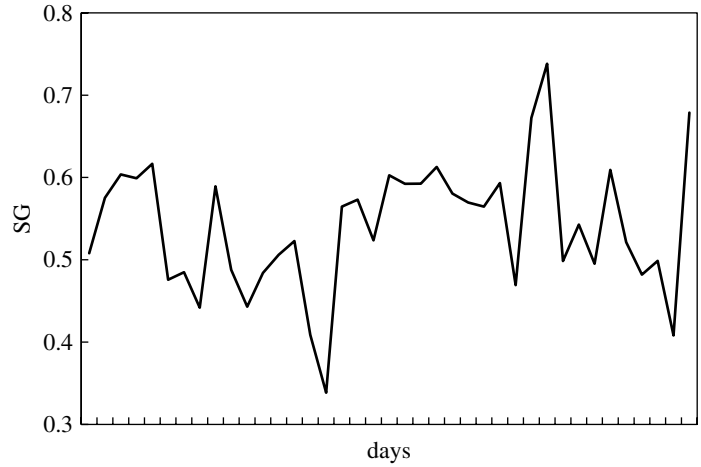

Figure 10.7 Variation of fuel gas SG

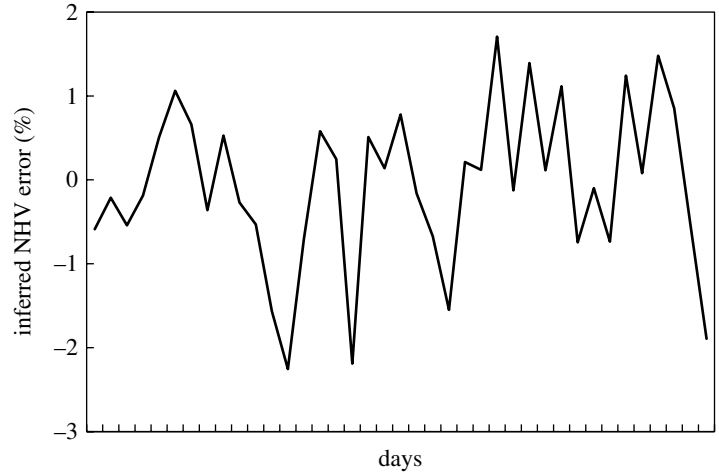

Figure 10.8 NHV prediction error

considerably less impact than the disturbances arising if there were no flow meter compensation. Figure 10.9 shows that much of the error is due to the varying inert content of the fuel gas.

From Equation (10.6) we can see that variation in SG has a linear impact on fired duty, rather than the square root relationship for orifice flow meters. Applying SG compensation to a flow meter measuring actual volumetric flow is therefore more important than applying it to an orifice flow meter. Conversely, for a coriolis flow meter, Equation (10.7) might indicate compensation is much less important. Indeed, this would be equivalent to making the assumption that heating value (on a weight basis) does not change with gas composition. However Figure 10.10 shows that, in this example, this is not the case. This is explained by Figure 10.11 which shows that most of the variation is due to changes in the hydrogen content – a gas which has (on a weight basis) a heating value more than double that of the lightest hydrocarbon. Varying over the range 40 to 70 vol% (8 to 15 wt%), the hydrogen content causes changes in NHV of around ±5 MJ/kg disturbing the fired duty of about ±10%.

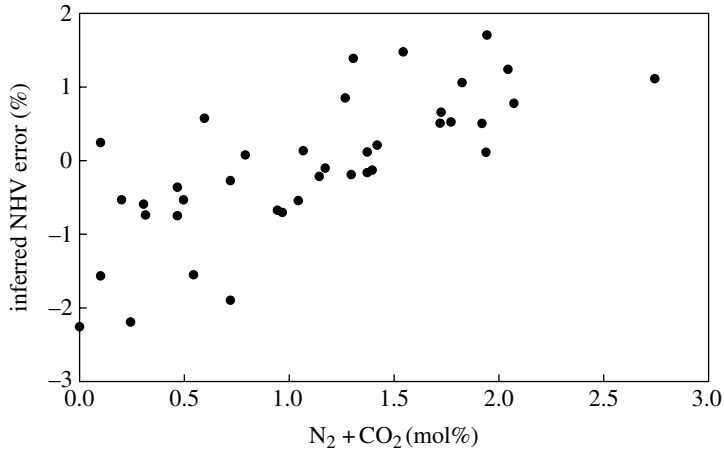

Figure 10.9 *Prime cause of NHV prediction error*

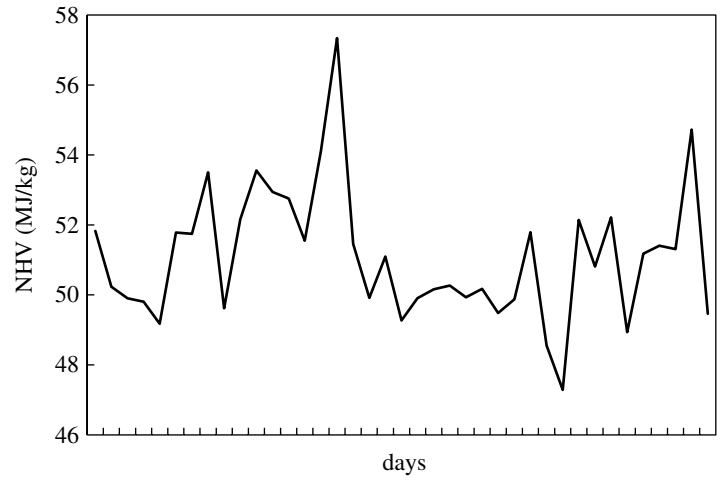

Figure 10.10 *Variation in NHV (weight basis)*

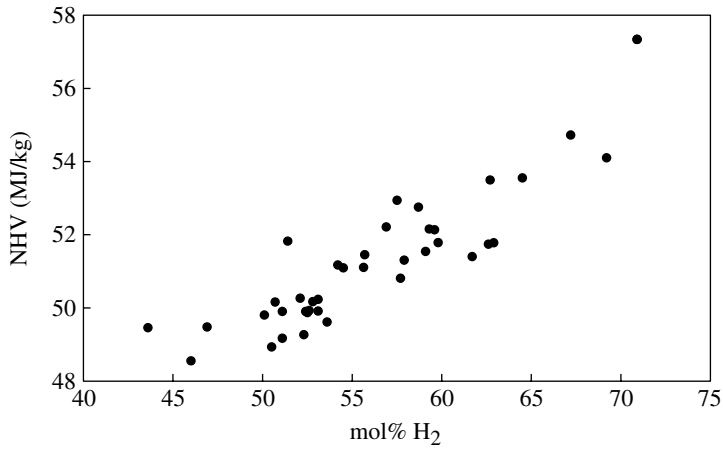

Figure 10.11 *Prime cause of NHV variation*

To determine what impact the use of uncompensated coriolis meters might have, Figure 10.12 presents the same data, but plotted as a frequency diagram. It shows, for example, that if our 39 laboratory samples were representative of a year's operation, there would be around 20 occasions per year where the daily disturbance to fired duty would be greater than 10%. Again, if such disturbances occurred slowly over a 24 hour period, this would not be of great concern. However, more rapid changes could present a major problem. This is particularly true if a site has adopted coriolis meters for all its heaters and boilers that share the same fuel gas supply. Simultaneously changing the duty on every process by 10% could cause a major disruption across the site.

10.2 Measuring NHV

The technique described above begs the question, why infer the NHV from SG instead of measuring it directly? Firstly, if we are using a pressure differential type flow meter, we need the SG measurement in any case for flow correction. Secondly, there is a large cost advantage. Densitometers are a much lower cost instrument than *calorimeters* or others that can be used, such as chromatographs. Further their installation costs considerably less. They are mounted on the pipework itself, much like a flow meter, and do not require an analyser house. Finally, there is a dynamic advantage. The residence time in the fuel gas system can be just a few seconds. Densitometers give an almost immediate indication of any change. Any delay could result in the heater outlet temperature controller (or steam pressure controller) detecting the disturbance, and taking corrective action, before the analyser responds. The now belated correction for the composition change would then cause a second disturbance.

We similarly need to ensure that the composition change is not dealt with too early. This can occur if the fuel gas supply pipework is long and the analyser located well upstream of the heater. Before embarking on analyser installation it is important to calculate the residence time between the proposed sample point and the heater. If this is significant then it is possible to delay the measurement in the DCS by the use of a deadtime algorithm. However, in this case, the variability of the residence time should also be checked. If the supply is dedicated to the heater, then this involves simply checking the maximum and minimum firing rates for that heater. However, if there are several heaters on site, it is common for there to be a fuel

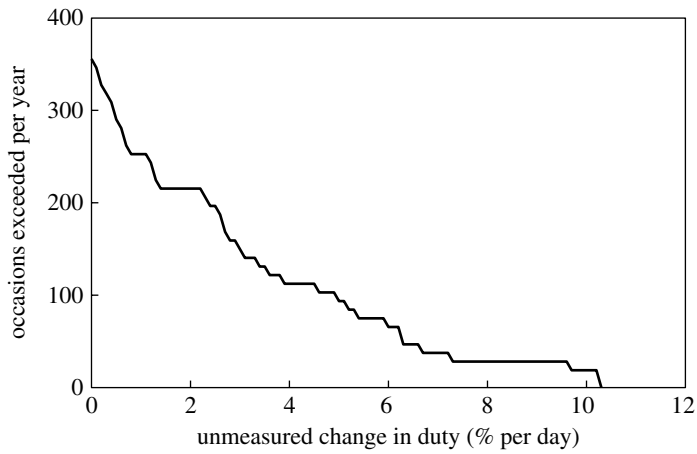

Figure 10.12 *Frequency of process disturbances*

gas header supplying all of them. The impact that the demands of the other heaters has on residence time then needs to be taken into account.

Theoretically, it is possible to automatically adapt the tuning of the deadtime algorithm based on measured gas flows but this is complex and prone to error if there are multiple fuel gas consumers and producers. It is likely to be more practical to locate the analyser close enough to the heater so that dynamic compensation is not required.

If the residence times for all the heaters are short, then it may be possible to locate the analyser on the shared header so that it may be used in all the firing controls. Under such circumstances greater attention should be given to the integrity of the whole system. A failure which causes the analyser to generate a low, but still believable, measurement would cause the fuel gas consumed by all heaters and boilers to rise simultaneously – potentially causing a major pressure disturbance in the fuel header. Given the relatively low cost of densitometers, it is practical to install two close together and cross-check their measurements. Any significant difference between their readings causes the scheme to switch to use the last good value rather than the current value. This provides for graceful degradation of the controller. However, when the fault is cleared, each controller using the value should be reinitialised. The new reading could be quite different from the last good reading and re-commissioning would otherwise 'bump' the process.

The choice of densitometer is important. Firstly, in order to avoid any measurement delay, it should be of the probe type and not one involving a sample withdrawal system. Secondly, remembering we are using it to infer MW, it should measure density at standard conditions not at stream conditions. While we can of course convert from one to the other in the DCS; this requires measurement of temperature and pressure at the sample point. It is more cost effective for this to be done within the analyser.

If the fuel contains significant concentrations of gases, other than hydrogen and hydrocarbons, whose concentration can vary quickly, then the inferred NHV may be unreliable. For example, if there was a large increase in the N_2 or CO_2 content of the gas in Table 10.2, the SG would increase but the NHV would reduce. The correction proposed in Equations (10.5), (10.6) and (10.7) would therefore cause the flow controller to reduce gas flow, when we need it to increase it. Such a problem would have been apparent when developing the correlation. A possible solution is the use of a *Wobbe Index (WI)* analyser, where

$$WI = \frac{NHV}{\sqrt{SG}} \tag{10.8}$$

The reasoning behind this choice of parameter is historical. Under the right circumstances, it will be a measure of fired duty. Flow (F) through a circular restriction of diameter (d), with discharge coefficient (c_d) is governed by an equation of the form

$$F = c_d \frac{\pi d^2}{4} \sqrt{\frac{2dp}{\rho}} \tag{10.9}$$

We discuss later in this chapter whether control of burner inlet pressure is appropriate but, if such a scheme is in place then, since firebox pressure is approximately constant, the pressure drop (dp) across the burner will then be constant. We have seen in Chapter 5 that temperature variation can normally be ignored and so gas density (ρ) will be proportional to SG. The flow through the burner will then be inversely proportional to $\sqrt{SG}$ and so fired duty will be proportional to WI. However, WI can also be used to determine fired duty when the gas flow is measured directly by an orifice type meter or similar instrument.

The analyser measures the NHV (on a volumetric basis) by continuously withdrawing and burning a sample of the gas. One approach is to record the amount of air necessary to consume it fully. The air flow

is adjusted to maintain a constant temperature of the combustion products. Figure 10.13 shows that there is strong correlation between the minimum air requirement and the NHV of common fuel gas components – including inerts and non-hydrocarbons. While not perfect, and not measuring the minimum air required, the analyser can be calibrated to give a measurement accurate enough for control. Another approach is to fix the air flow and then measure the residual oxygen in the combustion products. Both approaches also include a measurement of SG for use in the calculation of *WI*. Using *WI*, the energy flow calculation for a pressure differential type flow meter becomes

$$F_{true} = F_{measured}\sqrt{SG_{cal} \times \frac{P}{P_{cal}} \times \frac{T_{cal}}{T} \times WI} \tag{10.10}$$

Wobbe Index analysers are slower than densitometers but do give a continuous measurement. Whether the measurement is fast enough, however, should be checked on a per case basis.

Another possible type of analyser is a chromatograph. This would provide a full analysis of the fuel gas. Using known molecular weights and heating values for the components it is possible to accurately calculate the combined properties. Many chromatographs support this feature or the calculations could be located in the DCS. However, a full component analysis would take several minutes, by which time the feedback controller will have already corrected for the disturbance. In this case a chromatograph would provide an accurate measurement suitable for accounting and monitoring purposes but would be entirely unsuitable for control.

10.3 Dual Firing

Dual fuel firing was cited in Chapter 6 as an example of bias feedforward control. Here we will expand on this technique. The general control problem is illustrated in Figure 10.14. In this example fuel A is a gas over which we have no control. Fuel B is a liquid and its flow may be manipulated to control the heater outlet temperature. The scheme includes a bias feedforward scheme so that changes in fuel A are immediately compensated for by adjusting the flow of fuel B. For this to succeed we have to convert the units of

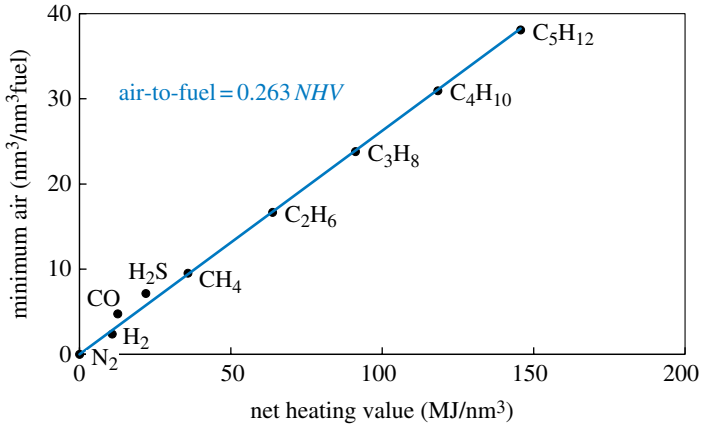

Figure 10.13 *Stoichiometric air requirement*

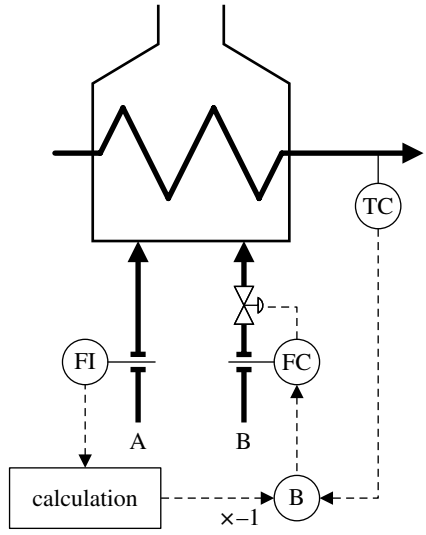

Figure 10.14 Dual firing example

measure of fuel A to be consistent with those of fuel B. By including the heating value of fuel B (NHV_b) in Equation (10.5) we get

$$F_a = F_{measured}\sqrt{\frac{SG_{cal}}{SG}\times\frac{P}{P_{cal}}\times\frac{T_{cal}}{T}\times\frac{(aSG+b)}{NHV_b}} \tag{10.11}$$

With the right choice of units the flow of fuel A (F_a) will now be in TFOE (tons of fuel oil equivalent). In the oil industry the barrel is commonly used as a measure of volume and so BFOE might be used. The output of the temperature controller can be thought of as the total duty demand (in fuel B units) from which is subtracted that delivered by fuel A.

10.4 Heater Inlet Temperature Feedforward

In Chapter 6 we used a fired heater as the example of an application of feedforward control on feed rate. The inclusion of fuel gas pressure, temperature and SG in the duty controller can also be thought of as feedforward schemes. Another potential source of disturbances is heater inlet temperature. Feed to the heater is often preheated by heat exchange with streams in other parts of the process. Any change in the flow or enthalpy of these streams can therefore cause a change in inlet temperature. The disturbance then caused to the outlet temperature is likely to be of a similar size; and the disturbance may be propagated again back to the heater inlet via the heat integration. Figure 10.15 shows the addition of a suitable bias feedforward scheme.

From Equation (6.23) we know that the gain in the bias feedforward is given by

$$K = -\frac{(K_p)_d}{(K_p)_m} \tag{10.12}$$

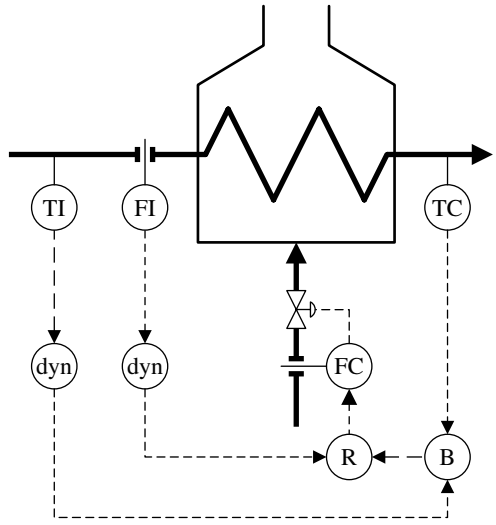

Figure 10.15 *Inlet temperature feedforward*

From Chapter 2, we know that $(K_p)_m$ varies inversely with feed rate. If we were to configure the output of the bias algorithm to manipulate the fuel flow directly, rather than the fuel-to-feed ratio, then we would need to include adaptive tuning to automatically adjust K to keep it in proportion to feed rate.

By definition

$$\left(K_p\right)_d = \frac{\Delta PV}{\Delta DV} \qquad \text{and} \qquad \left(K_p\right)_m = \frac{\Delta PV}{\Delta MV} \tag{10.13}$$

Thus

$$K = -\frac{\Delta MV}{\Delta DV} \tag{10.14}$$

For simplicity we assume that feed specific heat (c_p) and heater efficiency (η) are constant. If ΔT is the change in inlet temperature, then the required change in fuel flow (ΔF energy units) is given by the heat balance

$$F_{feed}.c_p.\Delta T = \Delta F.\eta \tag{10.15}$$

Combining with Equation (10.14) confirms the dependence of K on the feed flow (F_{feed}).

$$K = -\frac{\Delta MV}{\Delta DV} = -\frac{\Delta F}{\Delta T} = -\frac{F_{feed}c_p}{\eta} \tag{10.16}$$

This is not a problem, of course, if feed rate varies little. But if we can retain the feedforward ratio, the MV becomes the ratio SP. By dividing Equation (10.16) by F_{feed} we get

$$K = -\frac{\Delta R}{\Delta T} = -\frac{c_p}{\eta} \tag{10.17}$$

As it did with feedback control (Chapter 6) the use of the ratio obviates the need, as feed rate changes, to adjust the tuning of the feedforward scheme.

We may choose not to have the flow of the manipulated fuel to be conditioned to energy units if, for example, its properties change little. We then need to include the *NHV* in the calculation of *K*, ensuring we use engineering units consistent with the units of flow.

$$K = -\frac{\Delta R}{\Delta T} = -\frac{c_p}{\eta.NHV} \tag{10.18}$$

While we can predict the value for *K*, we still need to perform plant tests to obtain the process dynamics required for calculation of the tuning for the dynamic compensation. The procedure is covered in Chapter 6. We need to be careful with the units of *K*. It is usual for bias algorithms, unlike PID controllers, to operate in engineering units and so *K* should be determined in engineering units. Some model identification packages, however, work in fraction of range. If this is the case, if *K* is determined using Equation (10.12), it should be converted to engineering units by multiplying by

$$\frac{MV_{range}}{DV_{range}} \tag{10.19}$$

The result can then be checked for consistency with that obtained from Equation (10.17) or (10.18).

On some heaters, performing the necessary step tests may be impractical. We need to change the inlet temperature with the outlet temperature controller in manual mode. Introducing disturbances to the inlet temperature may not be straightforward and it would be inadvisable to leave the outlet temperature controller in manual mode for long periods waiting for a natural disturbance. A better approach would be to commission the feedforward controller with the value derived theoretically for *K* and dynamic compensation estimated assuming the process dynamics of inlet temperature changes are the same as those for feed rate changes. Careful monitoring of control performance would allow the tuning of the dynamic compensation to be modified as necessary.

An approach sometimes documented is to use the heat balance to predict the flow of fuel required.

$$F_{fuel} = \frac{F_{feed} c_p \left(T_{out} - T_{in}\right)}{\eta.NHV} \tag{10.20}$$

It is suggested that the result of this calculation is used as the SP of the fuel flow controller. Any feedforward scheme must have a feedback element; this is achieved by the feedback controller applying a bias to the predicted fuel requirement. While first appearing sound in principle, this approach does have problems.

- It is often implemented incorrectly. We have to choose whether to use the PV or the SP for T_{out}. Using the PV will mean that fuel is wrongly increased if the heater temperature rises. Using the SP resolves this but tuning the feedback controller then becomes difficult because both it and the feedforward controller will then respond to SP changes. Some handle this problem by lagging the SP used by the calculation. However, the proper approach is to use a fixed value for T_{out} so that incremental changes are

added, to the output of an otherwise conventional temperature-to-fuel flow cascade controller output, only if F_{feed} or T_{in} changes.

- The scheme fails to exploit the key benefit of ratio feedforward (as described in Chapter 6) in that it will be necessary to retune the temperature controller if the heater can operate over a wide range of feed rates.
- The scheme is unreliable if partial vaporisation takes place. This absorbs the majority of the energy supplied so that, if the degree of vaporisation changes, the energy requirement cannot accurately be predicted.

10.5 Fuel Pressure Control

Some fired heaters have a pressure, rather than flow, controller on the fuel – as shown in Figure 10.16. This is thought to give improved control if fuel gas composition varies. From Equation (10.9), fuel flow (F) through a single burner depends on the pressure drop across the burner (dp) and gas density (ρ).

$$F \propto \sqrt{\frac{dp}{\rho}} \tag{10.21}$$

For a constant firebox pressure, the pressure controller holds dp constant. An increase in gas density will therefore result in a reduction in gas flow. Since an increase in density usually reflects an increase in the NHV of the gas, directionally this behaviour is beneficial. However, flow should change in inverse proportion to NHV, not in inverse proportion to its square root.

Pressure control might also be considered preferable because, by clamping the SP of the pressure controller, burner pressure is kept within limits. Too high a pressure can result in the fuel velocity exceeding the flame velocity so that the flame separates from the burner tip and is potentially extinguished. Too low a pressure on oil burners can result in poor atomisation of the fuel and thus poor combustion. If the high SP limit is reached the operator will be expected to put additional burners into service, and to take some

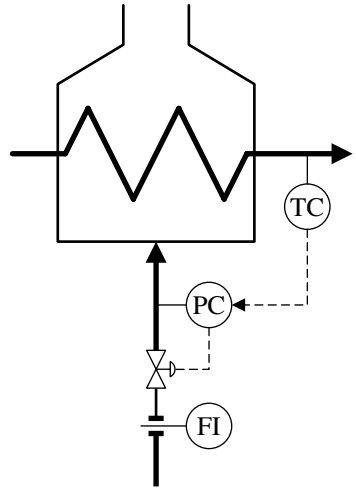

Figure 10.16 *Fuel pressure control*

out of service if the low pressure limit is reached. Under these circumstances, pressure control is worse than no control. For example, shutting the supply to a burner will cause an increase in pressure to the other burners and so the pressure controller will begin to close its valve. However, with fewer burners in service, we need a higher burner pressure to maintain the same duty and so we require the pressure control valve to open. The feedback controller will achieve this by increasing the SP of the pressure controller but has to recover from the disturbance caused by the valve initially being moved in the wrong direction. Indeed, the overall performance is likely to be improved by switching the feedback controller to directly manipulate the valve rather than cascade it to a pressure controller. On a baseload boiler there is usually no feedback controller and so steam production will be shifted, perhaps undesirably, to the swing boiler.

Another problem created by pressure control can be variation in the process gain between the heater outlet temperature (or header pressure) and the fuel pressure SP. The flow of fuel depends not only on this pressure but also on any resistance downstream of its measurement. So, for example, where burners are subject to fouling the relationship between fired duty and pressure will change over time. Changing the number of burners in service has a similar effect which, in the case of the need to take fouled burners out of service for cleaning, will further aggravate the problem. This is illustrated by Figure 10.17. With one burner in service, increasing the fuel pressure from 2 to 3 barg will increase fuel flow by about 15. With five burners in service the increase in fuel will be five times larger, so therefore will be the increase in outlet temperature (or steam production).

Pressure control also prevents the application of many of the techniques covered in this chapter. It would be difficult to devise a feedforward scheme to deal with disturbances in feed rate, heater inlet temperature or fuel heating value. Flow control, as we have seen previously in this chapter, can be enhanced to properly respond to changes in gas density. This technique provides precise control of fired duty whereas that provided by pressure control is only directionally correct.

One might consider cascading a flow controller to the pressure controller, rather than replacing it. This would allow retention of the pressure SP clamps and would permit the implementation of higher level schemes. The heater would still suffer the problem of the fuel control valve initially being moved in the wrong direction when the number of burners is changed. The scheme might also present a tuning problem. Controllers should not normally be cascaded unless the dynamics of the secondary are significantly faster

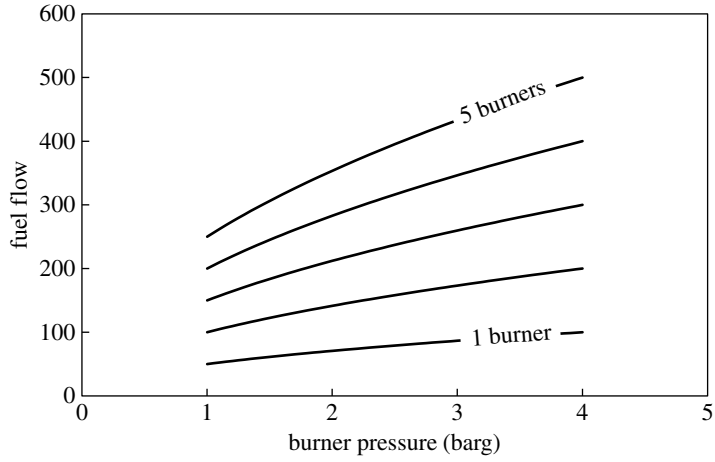

Figure 10.17 *Effect of the number of burners in service*

than those of the primary. In this application they would be virtually identical. Having to detune the flow controller to retain stability would undermine its advantage.

A better solution is illustrated in Figure 10.18. This has both a flow controller and a pressure controller. The flow controller permits the addition of the compensation and feedforward schemes. The pressure controller provides burner protection. In this case their outputs are compared in a low signal selector. This provides protection against too high a pressure. If the pressure exceeds the maximum (as entered as the SP in the pressure controller) the controller output will reduce to close the valve and so override the flow controller. The temperature will then not reach its SP and, as usual, the operator will need to put additional burners into service to relieve the constraint.

If protection against too low a pressure is required then a high signal selector is used. If protection against both low and high pressures is required then many DCS support a middle signal selector. If not, then low and high signal selectors can be configured in series as shown in Figure 10.19.

Another scheme which makes duty control difficult is that often installed on high viscosity fuel oil systems. To prevent pipework blockages, such fuel needs to be kept above a minimum temperature. Should its flow drop, heat losses from the pipework can result in the temperature falling below this minimum. To ensure a flow is maintained, even if a heater is shutdown, fuel is circulated around the site via a heated storage tank. The pipework passes alongside every heater and each burner on the heater can have its own take-off. Since it is not practical to measure the flow to an individual burner, flow meters are installed on the supply to and return from the heater. Fuel consumption is then determined by the difference between these measurements. However, because consumption is small compared to the circulating flow, the calculation is very prone to measurement error. For example, if the supply flow is 100%, measured to ±2%, and the return flow is 95%, also measured to ±2%, then the calculated consumption could vary by a factor of nine, i.e. from 1% to 9%. The fuel flow derived in this way might be considered as a feedforward DV for use in a total duty controller. Provided any measurement error is a sustained bias, then this would be feasible. However, if the error is random, then its use is likely to cause unnecessary correction and so worsen control of the heater.

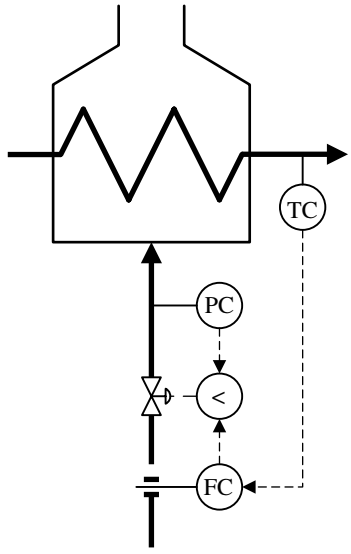

Figure 10.18 *Fuel high pressure override*

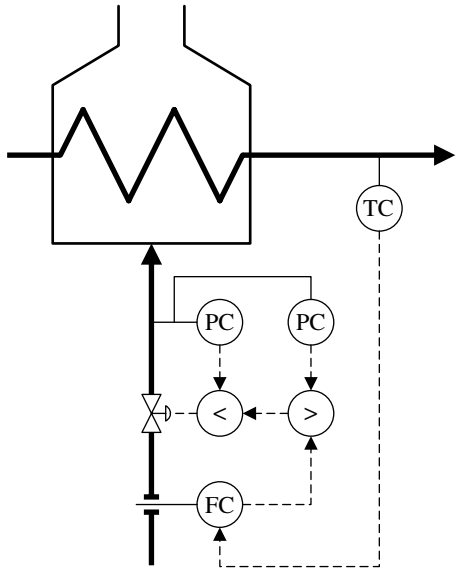

Figure 10.19 *Fuel low and high pressure override*

10.6 Firebox Pressure

In order to avoid the risk of flame exiting through inspection openings, etc. the pressure in the firebox must be kept below atmospheric pressure. This occurs naturally due to the difference in static head between inside and outside the firebox. If h is the height, ρ the density and g the acceleration due to gravity (9.81 m/s^2) then, in general, static head (P) is defined as:

$$P = h\rho g \tag{10.22}$$

Density can be derived from molecular weight (MW), absolute pressure (P) and absolute temperature (T) by applying the Ideal Gas Law to give:

$$\rho = \frac{P.MW}{RT} \tag{10.23}$$

R is the Universal Gas Constant (8314 J/kg-mole/K). For simplicity, we assume that P is atmospheric pressure (P_a). The difference (ΔP) in static head between inside and outside the firebox is therefore given by:

$$\Delta P = \frac{hP_a g}{R}\left(\frac{MW_{flue}}{T_{flue}} - \frac{MW_{air}}{T_{air}}\right) \tag{10.24}$$

We assume that ambient temperature (T_{air}) is 15°C (288.15 K) and that the flue gas temperature (T_{flue}) is 600°C (873.15 K). The units of ΔP will be the same as those of P_a. Firebox pressure, also known as *draught*, is very small and so typically measured in mm H_2O. Atmospheric pressure is 1.01325 barg or 10332 mm H_2O.

Consider the combustion of a fuel such as methane (CH_4). From Table 10.1 we see that air is 20.95% O_2. For simplicity let us assume the remainder is N_2.

$$CH_4 + 2O_2 + 2\left(\frac{79.05}{20.95}\right)N_2 \rightarrow CO_2 + 2H_2O + 2\left(\frac{79.05}{20.95}\right)N_2 \tag{10.25}$$

From Table 10.1, MW_{air} is 28.96. The molecular weight of the flue gas is given by:

$$MW_{flue} = \frac{44.01 + (2 \times 18.02) + \left(2 \times \frac{79.05}{20.95} \times 28.01\right)}{1 + 2 + \left(2 \times \frac{79.05}{20.95}\right)} = 27.63 \tag{10.26}$$

If we assume the *draught gauge* is mounted 3 m (h) above the air inlet, Equation (10.24) gives a result of −2.5 mm H_2O. This is a typical value. Because of the turbulence inside the firebox it can be a very noisy measurement.

10.7 Combustion Air Control

Figure 10.20 is a schematic of a typical combustion air system. Ambient air is routed to the *air preheater* by the *forced draught fan*. Control of air flow may be means of a variable speed drive on the forced draught fan or by some form of throttling of fan suction or discharge. Measurement of flow usually is by pitot tube or annubar. Air then divides to each of the burners; adjustment to individual burners is performed manually by manipulation of *air registers*.

The *stack damper* is adjusted manually to ensure the firebox pressure remains negative. The *preheater bypass* will generally be closed but can be opened as necessary should the flue gas temperature fall below dew-point. Any water condensed is likely to be corrosive due to acid gas products produced by the combustion of trace amounts of sulphur compounds such as H_2S. The *induced draught fan* returns the flue gases to the stack.

Alternative configurations are possible which omit either of the fans. If both fans are omitted, then recovery of heat from the flue gas is not possible using the type of air preheater shown. Control of air flow on such *natural draught* heaters is by manipulation of the stack damper, although only if this does not violate the need to keep firebox pressure negative.

Our control strategy must clearly ensure sufficient air is supplied for complete combustion of the fuel. Leaving unburnt fuel in the flue gas is potentially hazardous. Remembering that the firebox pressure is negative, any *tramp air* which enters through leaks above the combustion zone could potentially cause an explosion in the convection section. Incomplete combustion can result in soot particles thus producing black smoke with the resulting impact on the environment and fouling of any flue gas heat recovery or treatment systems. Once initiated, the problems associated with incomplete combustion can escalate quickly. Because the efficiency drops sharply the heater outlet temperature (or steam pressure) will fall; the controller will then respond by increasing the fuel flow.

However, it is undesirable to operate with excess air. Since air enters the heater at ambient conditions and leaves at flue gas temperature any unnecessary air increases the fuel requirement. Ideally we would like to maintain a stoichiometric air-to-fuel mixture. In practice we need to provide a small amount of excess air to compensate for incomplete mixing and to ensure that full combustion has taken place before the products of combustion leave the firebox.

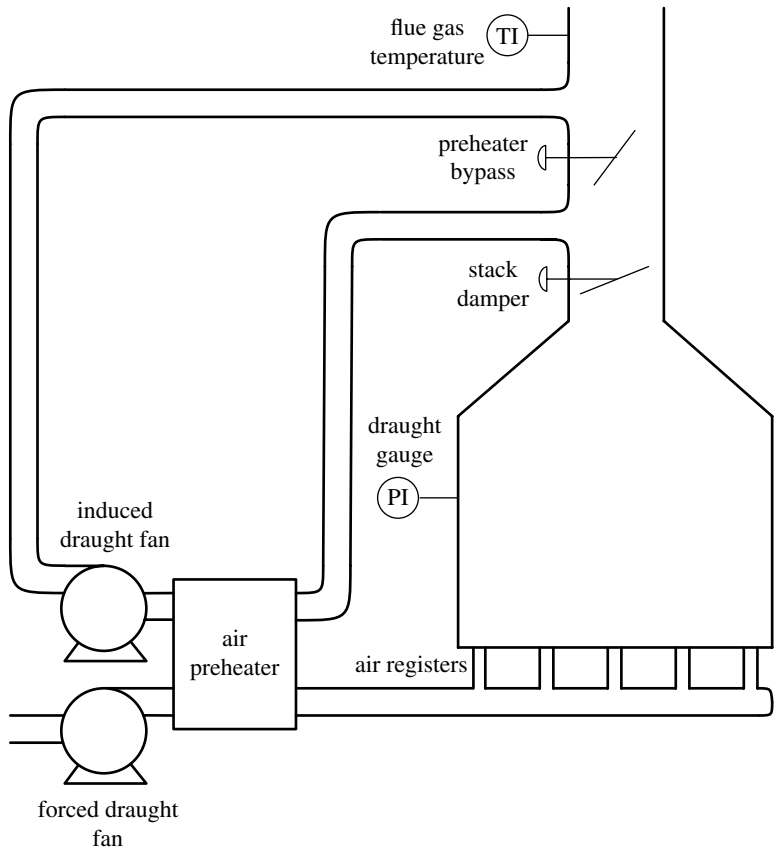

Figure 10.20 *Fired heater combustion air schematic*

Figure 10.21 shows the effect that varying the excess air has on flue gas composition, using methane (CH_4) as the example. By measuring flue gas composition we can assess the level of excess air. The most practical measurement is O_2. As with fuel gas properties we need an analyser that responds quickly. Oxygen analysers are available as probe types that can be inserted directly into the stack. Provided there is excess air, oxygen varies approximately linearly with air flow making tuning of the controller straightforward.

The relationship between O_2 and excess air can be developed by examining the chemical equations of combustion. If we first consider pure hydrogen, then

$$H_2 + 0.5O_2 \rightarrow H_2O \tag{10.27}$$

Then, for 10% excess air:

$$H_2 + 0.55O_2 + 0.55\left(\frac{79.05}{20.95}\right)N_2 \rightarrow H_2O + 0.05O_2 + 0.55\left(\frac{79.05}{20.95}\right)N_2 \tag{10.28}$$

On a *dry basis*, where the water remains as a vapour, the molar concentration of O_2 in the flue gas is given by

$$100 \times \frac{0.05}{1 + 0.05 + 0.55\left(\frac{79.05}{20.95}\right)} = 1.60\% \tag{10.29}$$

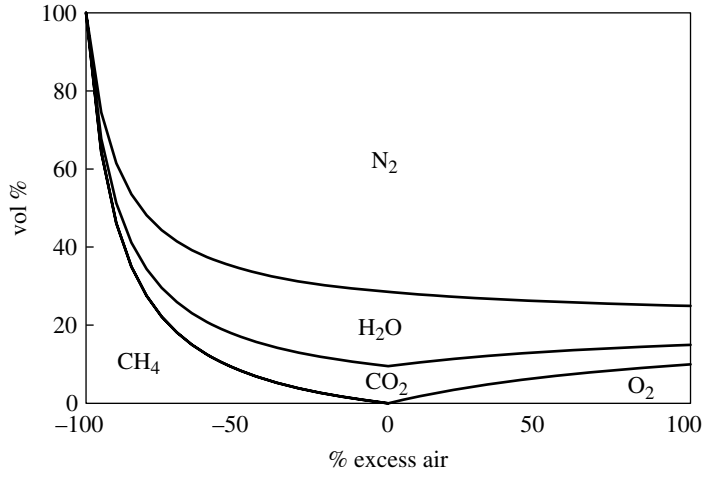

Figure 10.21 *Flue gas composition for combustion of methane*

On a *wet basis*, where the water is condensed to liquid, it becomes

$$100 \times \frac{0.05}{0.05 + 0.55\left(\frac{79.05}{20.95}\right)} = 2.35\% \tag{10.30}$$

We can repeat this exercise for pure carbon

$$C + O_2 \rightarrow CO_2 \tag{10.31}$$

$$C + 1.1O_2 + 1.1\left(\frac{79.05}{20.95}\right)N_2 \rightarrow CO_2 + 0.1O_2 + 1.1\left(\frac{79.05}{20.95}\right)N_2 \tag{10.32}$$

Since there is no water product, the wet and dry analyses are the same. The molar concentration of O_2 is thus given by

$$100 \times \frac{0.1}{1 + 0.1 + 1.1\left(\frac{79.05}{20.95}\right)} = 1.90\% \tag{10.33}$$

Figure 10.22 shows how the relationship between flue gas O_2 and excess air varies with fuel composition. This can be an issue on multi-fuel heaters. A low molecular weight fuel, such as a hydrogen/methane mixture, will have a H:C ratio of around 6. For a high molecular weight fuel, such as fuel oil, this ratio will be around 2. On a dry basis, switching from gas to oil will require operation at higher flue gas O_2 to give the same excess air. This, plus the fact that air-to-fuel mixing is less efficient with liquid fuels, is likely to require the operator to adjust the SP of the O_2 controller. Even with a fixed fuel composition, this is also likely to be necessary if the heater load varies because of changes in firebox residence time and the impact on mixing efficiency.

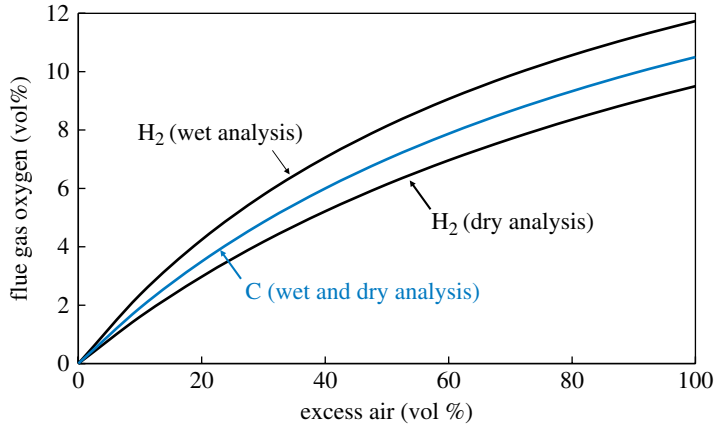

Figure 10.22 *Impact of fuel type on flue gas oxygen content*

The relationship between flue gas O_2 (x %) and excess air (y %), as described by Equations (10.30) and (10.33), can in general be written as:

$$x = \frac{20.95y}{a+y} \tag{10.34}$$

Rearranging

$$y = \frac{ax}{20.95-x} \tag{10.35}$$

The coefficient a depends on the fuel composition. For pure carbon it is 100 while for H_2 (and for CO) is 121. For natural gas (if pure CH_4) it is 110.5 and for fuel oil (and for H_2S) it is 107.

The measurement of flue gas O_2 theoretically offers another method of inferring NHV. By combining Equations (10.28) and (10.32), written in general for x % O_2, we can determine y % excess air from the molar H:C ratio (r).

$$a = 20.95\left(\frac{4+2r}{4+r}\right)+79.05 \tag{10.36}$$

So, if the fuel is close to pure carbon, r tends to zero and as expected Equation (10.36) gives a value for a of 100. Similarly, if the fuel is close to pure hydrogen, a is 121 because

$$r \to \infty \qquad \text{and} \qquad \frac{4+2r}{4+r} \to 2 \tag{10.37}$$

Equation (10.34) gives the dry analysis curves in Figure 10.22. To illustrate their use, if the flue gas O_2 is recorded as 4%, then from Equation (10.35) the excess air must lie between 23.6% (a = 100) and 28.5% (a = 121). However, fuel composition on a single heater will usually vary over a much smaller range.

For example, the laboratory results used to prepare Figure 10.6 show an average value of r of 2.65 with variation over the range 2.55 to 2.75. At 4% O_2 this variation would cause, in the estimate of excess air, an error of less than 0.08% of the true value. In theory we have also to consider the impact of non-hydrocarbon fuel types (such as CO and H_2S) but, as in this case, these are often in small enough quantities to be neglected.

Provided we have an accurate measure of the air flow at standard conditions (F_{air}), then we can use the value derived for y to determine the stoichiometric air requirement ($\hat{F}_{air}$).

$$\hat{F}_{air} = \frac{100F_{air}}{100 + y} \tag{10.38}$$

From the correlation shown in Figure 10.13, using the value defined by Equation (10.1) for the fuel gas flow (F_{true})

$$NHV = \frac{\hat{F}_{air}}{0.263F_{true}} \tag{10.39}$$

Rearranging gives a measure of the fired duty

$$F_{energy} = F_{true}NHV = \frac{\hat{F}_{air}}{0.263} \tag{10.40}$$

It is unlikely that this can replace the technique described in Equation (10.5), as the measurement of a fired duty controller, and so obviate the need for a measurement of fuel gas SG. At the very least it would require an accurate, fast and reliable measurement of both flue gas O_2 and combustion air flow. However, it might be used, initially at least, as a means of monitoring heater performance and validating other measures of NHV. Applying it to dual-fired heaters is even more problematic because of the complexity added by taking into account the air required for the liquid portion of the fuel.

While measurement of O_2 allows us to ensure sufficient excess air is delivered, it gives no indication of how sub-stoichiometric the air-to-fuel ratio might be. Thus an O_2 controller will respond at the same speed no matter how large the air shortage. Not shown in Figure 10.21 is CO. As the air-to-fuel ratio approaches the stoichiometric mixture, small amounts of CO will be detectable in the flue gas. This will increase markedly as air rate falls below the minimum required. The CO measurement cannot be used standalone because, like O_2, it only indicates over part of the operating range – showing zero no matter how much excess air is supplied.

If CO and O_2 are both present in significant quantities, then this may indicate the presence of tramp air. The combustion is sub-stoichiometric and air is entering the heater after the combustion zone. It indicates the need to seal leaks in the heater casing. If tramp air is not the problem, then simultaneous high readings can be built into the analyser validity check.

Many heaters cannot normally operate at the level where CO is detectable, but for those operating at 1% O_2 or lower, CO can be used to condition the O_2 measurement.

$$PV = \mathrm{O}_2 - K.\mathrm{CO} \tag{10.41}$$

The coefficient (K) is determined so that the process gain remains approximately constant over the whole operating range. By step testing at higher air rates we can obtain

$$\left(K_p\right)_{\mathrm{O}_2} = \frac{\Delta \mathrm{O}_2}{\Delta\left(\text{air/fuel ratio}\right)} \tag{10.42}$$

By step testing (carefully) at minimum air rate we can obtain

$$\left(K_p\right)_{\mathrm{CO}} = \frac{\Delta \mathrm{CO}}{\Delta\left(\text{air / fuel ratio}\right)} \tag{10.43}$$

Both process gains should be in engineering units. The coefficient (K) is then given by

$$K = \frac{\left(K_p\right)_{\mathrm{O_2}}}{\left(K_p\right)_{\mathrm{CO}}} \tag{10.44}$$

Alternatively it may be possible to use historical data to plot the lines shown in Figure 10.23 and determine K from the slopes of these lines. Figure 10.24 shows how the conditioned measurement varies with excess air. We now have an approximately linear relationship. Importantly it extends over the whole operating range, giving apparently negative O_2 measurements when there is insufficient air. The unavoidable increase in process gain, when both analysers show a measurement, can be considered a benefit. It will cause the controller to take corrective action more quickly as sub-stoichiometry is approached. Indeed, as an enhancement, the value of K can be increased slightly so that the controller will respond even more quickly when CO is detected. A cautious approach is advisable to ensure the controller does not become unstable under these circumstances.

The mechanics of implementing automatic control of flue gas O_2 can be complex. There may be problems with the O_2 measurement itself. Poor mixing of the flue gas may mean the measurement is not representative. Tramp air will cause a false measurement of what is happening in the combustion zone. Some heaters have multiple cells with their flue gas routed to a common duct. An analyser located here would be of little use for control. However, the development of non-invasive analytical technologies, such as tuneable diode laser absorption spectroscopy (TDLAS), now permit measurement to be taken across the radiant section of the heater and be unaffected by poor mixing. While more often installed as part of a safety system, they can be used for control of excess air.

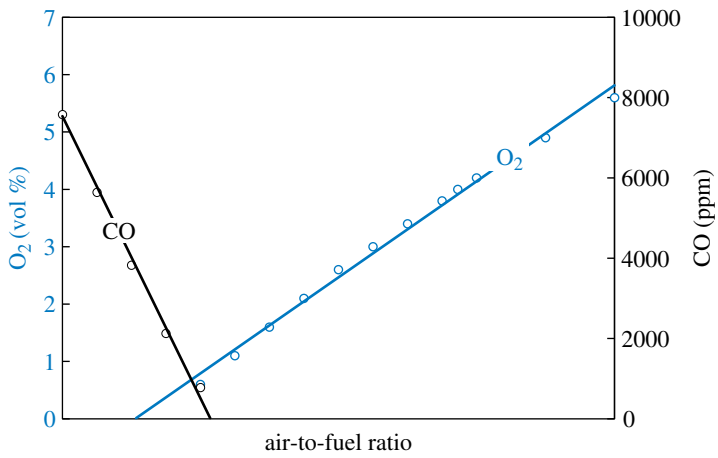

Figure 10.23 *Effect of air on flue gas analysis*

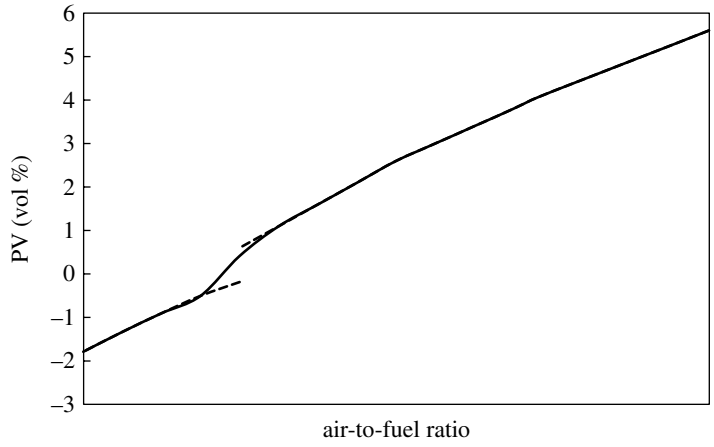

Figure 10.24 *Conditioned oxygen measurement*

There can be problems with manipulation of air flow. Natural draught heaters have no fan and so air flow is adjusted by manipulating the position of a stack damper. The relationship between air flow and damper position can be highly nonlinear. The damper is in a potentially corrosive and dirty environment and prone to mechanical failure. It can often exhibit stiction or hysteresis. On forced draught heaters there is the option of throttling the suction or discharge of the fan but this can have many of the problems common to stack dampers. Manipulating the speed of an electrically driven fan, e.g. by using a *variable frequency drive (VFD)*, is an expensive option and can also be nonlinear. There are also potential operating problems; we have to be sure the control does not increase the probability of sub-stoichiometric combustion or positive firebox pressure.

These are not necessarily reasons for not progressing with improved control. The point is made to demonstrate that implementation may be costly – particularly if retro-fitted to an existing heater. Before embarking on implementation we need to ensure the economic payback makes it worthwhile.

The savings will depend on flue gas temperature, fuel type and how close the operation already is to minimum excess air. If the temperature of the flue gas is reduced by the installation of some form of heat recovery system, then the fuel wasted by excess air will be reduced. Figure 10.25 shows the effect, using ethane (C_2H_6) as the fuel example. Figure 10.26 shows the effect that fuel type has on the potential savings if the flue gas temperature is 400°C (around 750°F). Estimating the benefit of reducing excess air should take into account the impact on any heat recovery system. Reducing excess air is likely to affect firebox temperature and will certainly change flue gas flow and its composition. These changes will alter the efficiency of heat recovery. It is likely that flue gas temperature falls as excess air is reduced and so any saving estimated using Figure 10.26 can be inaccurate. It is worth noting the penalty for going sub-stoichiometric. This wastes about 10 times more fuel than the equivalent amount of excess air. Should operation at lower O_2 levels increase the probability of this occurring, then this cost, combined will the resultant operational problems, could well exceed the annual benefits captured by the controller. Figure 10.27 combines the impact, on fuel wasted, of O_2 and flue gas temperature. Actually drawn for methane (CH_4), the spacing between the curves is similar for all hydrocarbons and so can be used to estimate the fuel saving from reducing O_2 or stack temperature on most heaters.

Figure 10.28 shows in principle how O_2 control might be added to our dual fired heater example. It assumes some form of flow control is feasible on combustion air, on this occasion via a variable speed drive on the fan. Rather than directly manipulating this flow an air-to-fuel ratio has been installed. This

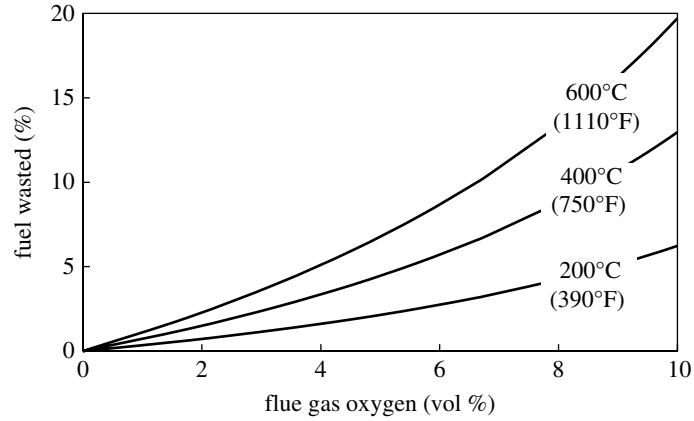

Figure 10.25 *Effect of flue gas temperature on fuel wastage*

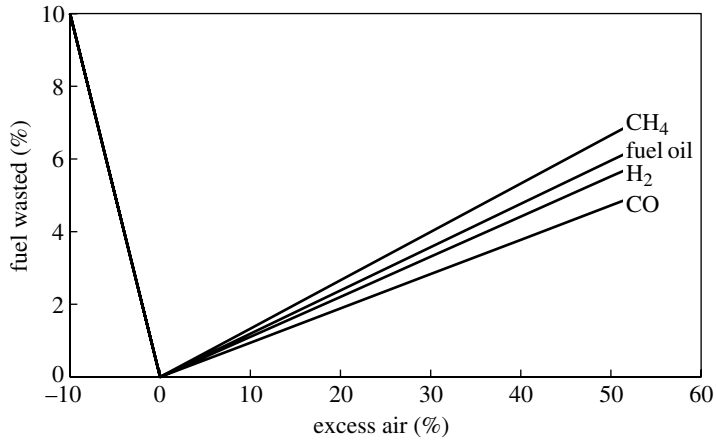

Figure 10.26 *Effect of fuel type on fuel wastage*

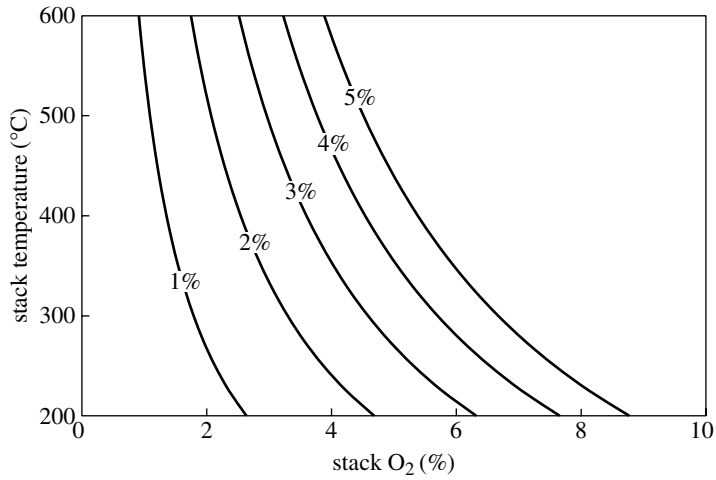

Figure 10.27 *Effect of flue gas oxygen on fuel wastage*

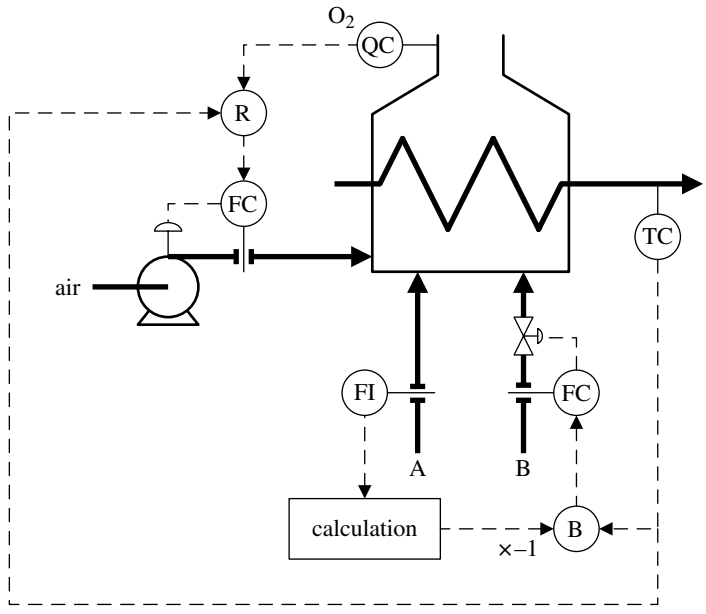

Figure 10.28 *Principle of flue gas oxygen control*

offers several advantages. Firstly, it will help maintain the excess air constant during times when the analyser is out of service. Secondly, as covered in Chapter 6, it will obviate the need to re-tune the controller as feed rate changes. This is particularly important on fired boilers which frequently have a turndown ratio of around 4:1. Without the ratio, a fourfold change in process gain would present a tuning problem. Thirdly, as we explain later in this chapter, its dynamics change when a boiler is switched between swing and baseload mode. The use of the air-to-fuel ratio stops this.

Some Wobbe Index analysers offer the option of also generating a measurement of air demand, known as the *Combustion Air Requirement Index (CARI)*.

$$CARI = \frac{\text{moles air/mole of fuel}}{\sqrt{SG}} \tag{10.45}$$

Close examination of Figure 10.13 shows that the minimum air-to-fuel ratio, for gases such as H_2S, H_2 and CO, deviates slightly from the relationship applicable to hydrocarbons (and inerts). As Figure 10.29 shows, this is predicted by *CARI*. If the content of these gases varies significantly, *CARI* could be used to adjust the SP of the air-to-fuel ratio. Or, if the SP is set by the stack O_2 controller, reset the minimum permissible ratio. The required air flow can be determined by replacing *WI* in Equation (10.10) with *CARI*.

In Figure 10.28 we used the fuel demand signal as our measure of fuel flow. One could argue that this is desirable on increasing demand since it is an earlier indication of the need to increase air flow than that given by the actual fuel flow measurement. However, on a decreasing fuel demand, this is not the case. The simple ratio also presupposes that there is adequate air for the additional fuel; if the fan is at maximum capacity the feedback controller would still increase fuel if required. The scheme also assumes that the dynamics of air flow control are the same as those of fuel flow. Due to the size of the actuator it is likely

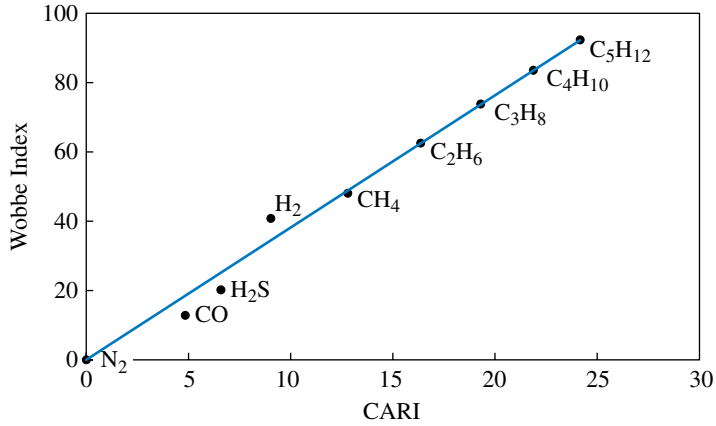

Figure 10.29 *Relationship between Wobbe Index and CARI*

that air flow will increase more slowly than fuel, possibly causing a transient shortage of air enough to take the mixture sub-stoichiometric.

To overcome these potential problems, a *cross-limiting* approach may be adopted. This is shown in Figure 10.30. It is also known as the *lead-lag* scheme (not to be confused with the lead-lag control algorithm). It gets its name because the air leads the fuel on increasing demand but lags it on decreasing demand.

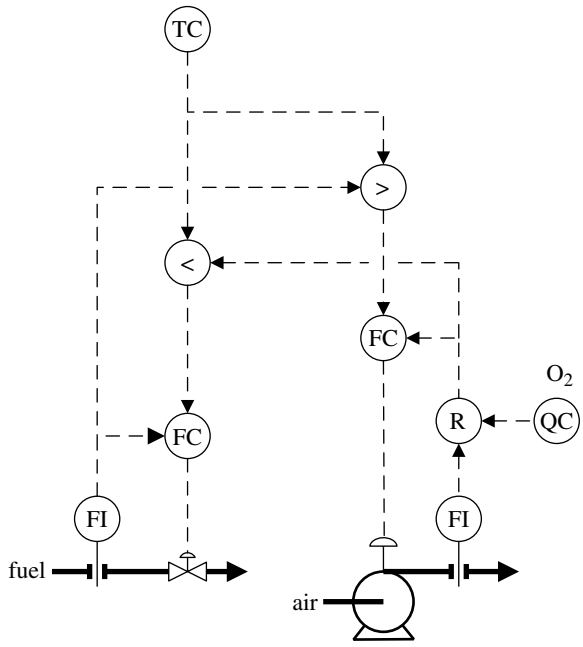

Figure 10.30 *Cross-limiting control*

The ratio is configured as fuel-to-air rather than the otherwise more usual air-to-fuel. It therefore converts the air flow measurement into units of fuel flow. The result is the flow of fuel that can safely be consumed at the current air rate. This value is used in two places. Firstly, it provides the measurement to the air flow controller. Secondly, it provides the input into a low signal override (<) on the output of the temperature controller. If the value is less than the fuel demand it prevents the fuel flow from being increased. However the fuel demand is also fed, through a high signal selector (>), as SP (in fuel flow units) to the air flow controller. Because it is increasing it is passed through by this selector. As the air flow controller responds to this increase, the override on the temperature controller is relaxed and the fuel SP is permitted to increase. Thus, on increasing demand, fuel will not be increased until sufficient combustion air is delivered.

Conversely, on decreasing demand, the temperature controller output is passed by the low signal selector to the SP of fuel flow controller. The fuel flow measurement provides an input to the high signal selector. If the fuel flow controller does not respond, then this will override the output sent to the air flow controller. Thus air is not permitted to reduce until the fuel flow reduces.

Some plant owners take the view that the complexity of the scheme creates a hazard because of potential misunderstanding of its operation by the process operator. One source of confusion is the use of the reciprocal of the air-to-fuel ratio. Another is the air flow controller operating in equivalent fuel units. It is possible to reverse the scheme so that the ratio is on the fuel side, which resolves these issues but then means that the fuel flow control operates in equivalent air units. It is also possible to reconfigure the scheme so that both ratios are used and both flow controllers work in their own units, as shown in Figure 10.31.

The scheme can give problems in tuning the TC (or PC). If the air flow controller is particularly sluggish then, on increasing fuel demand it will cause a significant process deadtime and/or lag. On decreasing fuel demand it has no effect. This nonlinearity means that the controller will not work well in both directions. The solution should be to first address the air flow controller to ensure it is mechanically sound and tuned to react as fast as possible. If the problem persists, then it will be necessary to obtain the dynamics for each

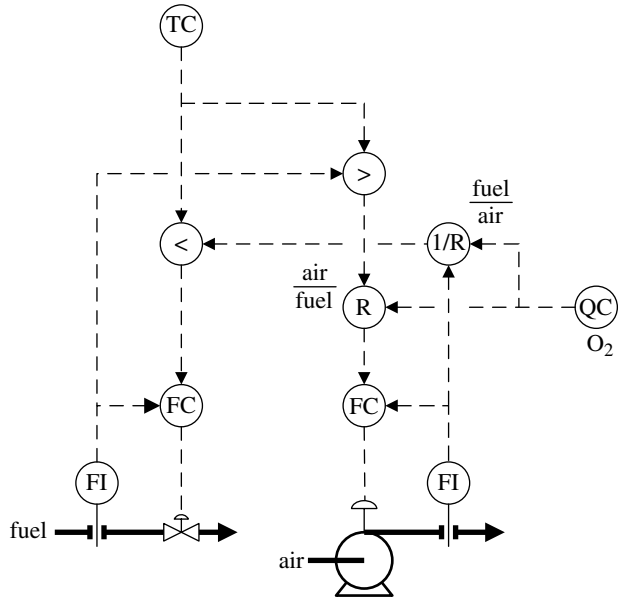

Figure 10.31 *Cross-limiting control (alternative)*

of the cases of air flow either over-riding the fuel flow or not. It should then be possible to develop some compromise tuning that works reasonably well in both directions. If not, then some form of adaptive tuning would be needed, which changes the controller tuning depending on whether its output is increasing or decreasing. Care also has to be taken in the design of scaling and initialisation.

Those concerned that the complexity can result in operators taking the wrong actions would cite situations such as the following. Imagine that the air flow controller has been switched to manual mode because of a mechanical fault. The temperature controller may not then be permitted to increase fuel when required. This will result in the temperature (or steam pressure) falling below SP. While arguably safer, the operator might then respond by also switching the fuel flow controller to manual and directly opening the fuel valve. This could potentially drive the combustion sub-stoichiometric. The correct approach would, of course, be to manually increase the air and let the control scheme increase fuel flow as required.

However, there are a large number of the schemes in place throughout industry operating successfully. Some plant owners mandate it as a standard, as do some professional bodies and some suppliers of fuel.

10.8 Boiler Control

Boiler control is similar to process heater control, except that boiler duty is manipulated by a steam header pressure controller. Of the schemes outlined so far in this chapter all would be applicable with the exception of the feedforward schemes on feed rate and inlet temperature. The flow of boiler feed water is changed in response to a change in steam demand or drum level, as we saw in Chapter 4, and so follows rather than leads an increase in firing. Feedforward on water flow therefore brings no dynamic advantage. It might be possible to ratio fuel to steam demand but the dynamic advantage is likely to be small since header pressure will respond quickly to any imbalance between supply and demand. It can also cause instability if the demand signal includes a measurement of the boiler's steam production. Any disturbance on the boiler will then be aggravated as the change in production causes unwanted feedforward action. The common advantage of ratio feedforward, i.e. not having the re-tune the controller as feed rate changes, does not apply to integrating processes. Steam header pressure control is such a process.

Whether a boiler is in baseload or swing mode affects its process dynamics. Thus, if an individual boiler can be switched between these modes, we must pay particular attention to the control design. For example, consider the result of a step test performed to help design flue gas oxygen control. The purpose of the test is to obtain the process dynamics of flue gas oxygen with respect to changes in air flow. If we change the air rate to a baseload boiler, then we change the efficiency of that boiler and hence its steam production. This will in turn disturb the header pressure and the pressure controller will take corrective action by adjusting the firing on the swing boiler. If we repeat the step test with the boiler now in swing mode, the pressure controller will change the duty of the boiler on which we are conducting the test. The change made to firing will affect the flue gas oxygen and we will therefore obtain a different result for the process gain. If we were to use the results of the first test to tune the controller, then there is a danger that it will become unstable when the boiler is switched to swing mode.

Of course we could automatically switch tuning with mode changes, but the air-to-fuel ratio scheme already proposed resolves the problem. With this ratio in place the result of testing in swing mode will be the same as that in baseload mode. In swing mode, where the header pressure control adjusts the duty, the air flow will be maintained in proportion and the change in flue gas oxygen will be unaffected.

If the boilers are of different designs, then the dynamic relationship between header pressure and boiler firing will vary – depending on which is selected as swing. Similarly the dynamics will change if the number of swing boilers is changed. Different tuning will then be required in the header pressure controller as its MV is changed.

10.9 Fired Heater Pass Balancing

A popular strategy is balancing passes within multi-pass fired heaters. Such heaters will have flow controllers on each pass. The concept is to keep the total flow through the heater at the required value but adjust the distribution of the flow between the passes. The simplest objective is to maintain equal pass flows. Probably the most common is to equalise the individual pass outlet temperatures in the belief that heater efficiency is maximised. Other strategies are possible if the heater is a bottleneck. For example, if it is the main hydraulic constraint, then equalising the positions of the pass flow control valves, so that no one flow controller saturates before the others, will maximise capacity. If the main constraint is tube skin (tube metal) temperature then, provided this can be measured reliably, these can be equalised. This would permit either the combined heater outlet temperature or the total feed rate to be increased. The run length of heaters in coking service is limited by the most coked pass. Balancing the rate of coking of each pass can be exploited by increasing the run length or operating at a higher severity.

In the case where we believe there are efficiency benefits we can predict these. The value of balancing pass outlet temperatures depends on the nonlinear relationship between fired duty and temperature. In the radiant section of the heater this is governed by *Stefan's Law* which states that rate of heat radiated is proportional to the fourth power of the absolute temperature. Assuming our heater operates with an outlet temperature of 300°C we can use this relationship to construct the curve in Figure 10.32. If the passes are imbalanced so that half are 10°C hotter than the mean, then our chart shows an additional 7.2% firing is required. For those passes 10°C below the mean, 6.8% less firing is required. So by exactly balancing the temperatures the overall saving will be 0.2% of total firing. Much of this saving might be achieved by occasional manual attention to the distribution but, even if not, it is unlikely that such an improvement would be measurable. If this is the only source of benefit, then it is unlikely to justify the cost of implementation.

However, pass temperature balancing can be very lucrative in situations where there is an economic incentive to operate at the highest possible combined heater outlet temperature. If limited by the metallurgy of the passes, then raising all the pass outlet temperatures to the limit would, in this example, allow the combined temperature to be increased by 10°C.

Balancing of heaters in coking service should be approached with care. A coked pass will be less efficient in transferring heat from the firebox to the fluid in the pass. Its pass outlet temperature will then be

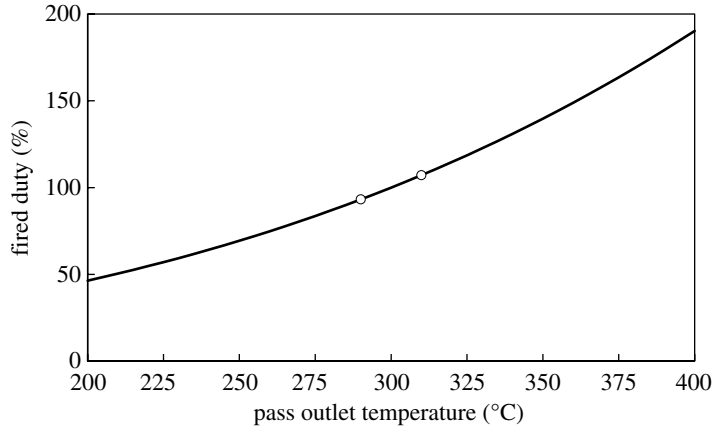

Figure 10.32 *Effect of pass outlet temperature on fuel demand*

lower. If the objective of pass balancing is to equalise pass temperatures, then it will reduce the flow through the pass in order to increase the temperature. This will increase the residence time and accelerate coking. Left unchecked, this will shorten run length.

If the heater is a hydraulic constraint, then we can use historical data to quantify the relationship between pass flow and valve position. By comparing average valve position to the maximum permitted we can then estimate the overall increase in capacity utilisation.

In the case of a coking constraint (and skin temperature constraint if caused by coking) we can perform a calculation similar to that for efficiency savings. Coking is a chemical reaction governed by the *Arrhenius Equation* which states that rate of reaction (k) is governed by the activation energy (E), the universal gas constant (R) and the absolute temperature (T).

$$k \propto e^{-E/RT} \tag{10.46}$$

The relationship is shown in Figure 10.33. By applying this to the same heater, those passes 10°C hotter than the average will coke 19.5% more quickly. Those 10°C cooler will coke 16.8% more slowly. We could interpret this as the potential to achieve a reduction of 1.3% in the overall rate of coking. And we could exploit this by increasing the heater outlet temperature by 0.7°C. Or we might exploit it by de-coking less often. However, our calculation assumes that the period that each pass operates above average temperature is the same as the time it spends below. If one pass is consistently the hottest, then it will limit run length. Balancing would then increase run length by 24%, i.e. 100/(100-19.5).

Of course we need some measurement of rate of coking for each pass. This might be skin temperature but these measurements are prone to failure. It may be possible to develop an inferential for rate of coking based on measurements such as pressure, temperature, flow and (if injected) steam flow.

Before embarking on pass balancing, it is advisable to check for interactions between pass flow controllers. Depending on the geometry of the pipework and the pressure differential, increasing the SP of one pass flow starves the others and their corrective action interacts with the first and causes an oscillatory response. The usual way of dealing with this is to configure different gains in each of the flow controllers.

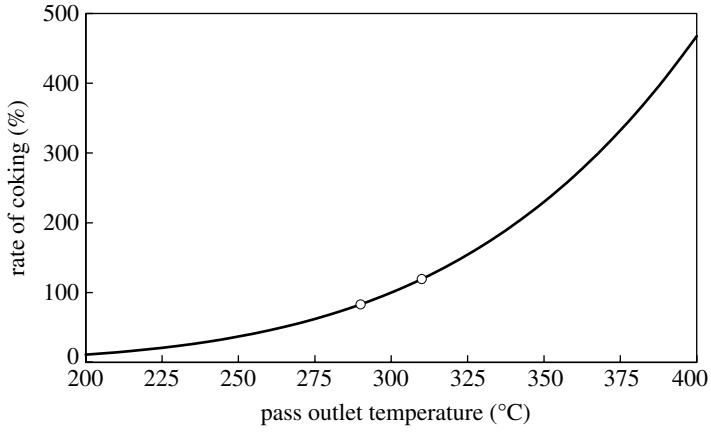

Figure 10.33 *Effect of pass outlet temperature on rate of coking*

Instrument calibration should also be checked. Often in a difficult service it is common for errors to occur in the flow measurements. For example, orifice tappings or the orifice itself may become partially blocked. An apparent maldistribution of flows may not be real – particularly if control valve positions (and/or pass outlet temperatures) are approximately equal.

In terms of the techniques available, there are three basic approaches. Although the schemes described are designed to balance temperatures they can be modified to meet any of the other objectives. The first approach is to use standard algorithms in the DCS as shown in Figure 10.34. Although, for simplicity, only a two-pass heater is illustrated the technique is applicable to any number of passes. One is selected as master. For each of the other passes the difference between its outlet temperature and that of the master is calculated. Each of these is the PV of a PID controller that has SP of 0. The master flow controller manipulates the master pass control valve directly and each of the others via a bias. Each bias is set by each of the temperature difference controllers. Upper and lower limits are set on the biases to prevent too large a flow imbalance. Indeed, since we have removed the individual pass flow controllers, it is important to monitor the flow imbalance and alarm any significant maldistribution. This might be an early sign of excessive coking in one pass.

An alternative approach is to custom code the technique. We define a predicted outlet temperature (T_{out}) based on each of the n pass flows (F_i) and outlet temperatures (T_i).

$$T_{out} = \frac{\sum_{i=1}^{n} F_i T_i}{\sum_{i=1}^{n} F_i} \tag{10.47}$$

Using the heater inlet temperature (T_{in}), we predict the pass flows (F_i^*) necessary to balance the heater.

$$F_i^* = \frac{T_i - T_{in}}{T_{out} - T_{in}} F_i \tag{10.48}$$

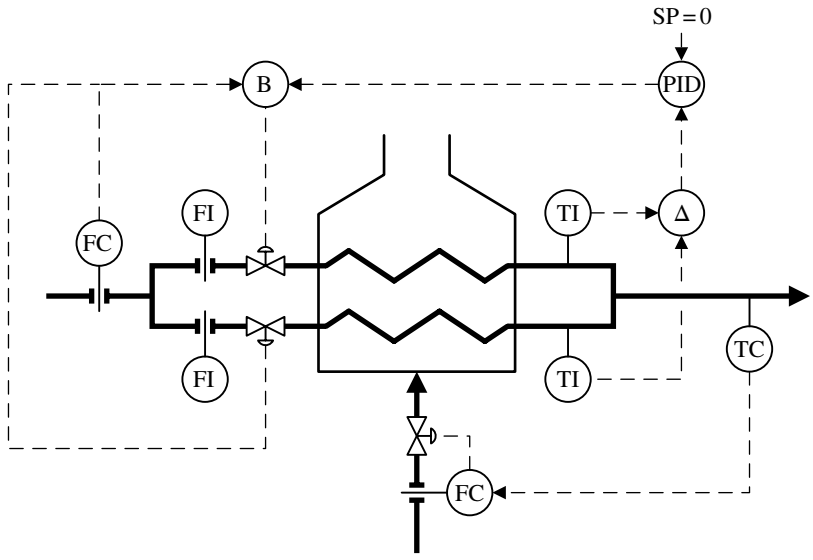

Figure 10.34 *Heater pass balancing*

The use of the derived outlet temperature rather than that measured ensures there is no change in the total flow through the heater.

$$\sum_{i=1}^{n} F_i^* = \sum_{i=1}^{n} F_i \tag{10.49}$$

The advantage of the custom code approach is that a wide range of other checks can be readily added and help improve process integrity. For example, validity checks can be made on all flows, temperatures and control valve positions. Constraints such as maximum flow imbalance, maximum skin temperature etc. can be included. Logic can be included to handle out of the ordinary situations, such as instrument failure, and the application degraded gracefully rather than an all-or-nothing approach.

The final approach is to make use of MPC. We can use the two-pass heater as a simple example. The manipulated variables MV_1 and MV_2 are the pass flows F_1 and F_2. The pass outlet temperatures are T_1 and $T_{2,}$ so from Equation (10.47)

$$T_{out} = \frac{F_1 T_1 + F_2 T_2}{F_1 + F_2} \tag{10.50}$$

We define CV_1 as the difference between the pass 1 outlet temperature and the combined outlet temperature (T_{out}).

$$CV_1 = T_1 - \frac{F_1 T_1 + F_2 T_2}{F_1 + F_2} = \frac{F_2 \left(T_1 - T_2\right)}{F_1 + F_2} \tag{10.51}$$

Similarly we define CV_2 as the difference between the pass 2 outlet temperature and the combined outlet temperature.

$$CV_2 = T_2 - \frac{F_1 T_1 + F_2 T_2}{F_1 + F_2} = \frac{F_1 \left(T_2 - T_1\right)}{F_1 + F_2} \tag{10.52}$$

The objective of the controller will be to reduce these differences to zero. Although the combined outlet temperature will likely be used for temperature control and so measured directly, CV_1 and CV_2 should still be derived from the calculated value. Otherwise measurement errors will make it impossible for the controller to make all the pass outlet temperatures exactly equal to the combined outlet temperature.

Let us assume that step-tests made to MV_1 gave the following process gains for CV_1 and CV_2:

$$\left(K_p\right)_{11} = \frac{\Delta CV_1}{\Delta MV_1} = -2.5 \qquad \text{and} \qquad \left(K_p\right)_{21} = \frac{\Delta CV_2}{\Delta MV_1} = 2.5 \tag{10.53}$$

And step-tests made to MV_2 gave:

$$\left(K_p\right)_{12} = 2.5 \qquad \text{and} \qquad \left(K_p\right)_{22} = -2.5 \tag{10.54}$$

To enable us to place some restriction on flow maldistribution, we define CV_3 as the difference between the flow through pass 1 and the average pass flow.

$$CV_3 = F_1 - \frac{F_1 + F_2}{2} = \frac{F_1 - F_2}{2} \tag{10.55}$$

And so

$$\left(K_p\right)_{31} = \frac{dCV_3}{dF_1} = 0.5 \quad \text{and} \quad \left(K_p\right)_{32} = \frac{dCV_3}{dF_2} = -0.5 \tag{10.56}$$

Similarly we define CV_4 as the difference between the flow through pass 2 and the average pass flow.

$$\left(K_p\right)_{41} = -0.5 \quad \text{and} \quad \left(K_p\right)_{42} = 0.5 \tag{10.57}$$

Finally we define CV_5 as the total flow through the heater

$$CV_5 = F_1 + F_2 \quad \text{and so} \quad \left(K_p\right)_{51} = \left(K_p\right)_{52} = 1.0 \tag{10.58}$$

The process can therefore be described by

$$\begin{pmatrix} \Delta CV_1 \\ \Delta CV_2 \\ \Delta CV_3 \\ \Delta CV_4 \\ \Delta CV_5 \end{pmatrix} = \begin{pmatrix} -2.5 & 2.5 \\ 2.5 & -2.5 \\ 0.5 & -0.5 \\ -0.5 & 0.5 \\ 1.0 & 1.0 \end{pmatrix} \begin{pmatrix} \Delta MV_1 \\ \Delta MV_2 \end{pmatrix} \tag{10.59}$$

As previously, pass 1 outlet temperature is 310°C and that of pass 2 is 290°C. The current operating strategy is to equalise the pass flows; assume each is currently 50 m^3/hr. We then commission the controller with the following constraints:

	LO	HI
CV_1	0	0
CV_2	0	0
CV_3	–3	3
CV_4	–3	3
CV_5	100	100

(10.60)

The limits on CV_1 and CV_2 are set to keep the pass temperatures exactly equal. Those on CV_3 and CV_4 permit the pass flows to deviate from the average by up to 3 m^3/hr. The limits on CV_5 are set to keep the total flow at 100 m^3/hr.

The heater is currently operating at the desired value for CV_5; thus we want ΔCV_5 to be zero. Therefore, from Equation (10.59)

$$0 = \Delta MV_1 + \Delta MV_2 \tag{10.61}$$

From Equation (10.51), CV_1 currently has a value of 10°C and so, to satisfy the LO/HI constraints, we require ΔCV_1 to be −10°C. From Equation (10.59)

$$-10 = -2.5\Delta MV_1 + 2.5\Delta MV_2 \tag{10.62}$$

Solving Equations (10.60) and (10.61) gives

$$\Delta MV_1 = 2 \quad \text{and} \quad \Delta MV_2 = -2 \tag{10.63}$$

Equation (10.59) shows what effect this will have on the other CVs.

$$\Delta CV_2 = \left(2.5 \times 2\right) + \left(-2.5 \times -2\right) = 10 \tag{10.64}$$

$$\Delta CV_3 = \left(0.5 \times 2\right) + \left(-0.5 \times -2\right) = 2 \tag{10.65}$$

$$\Delta CV_4 = \left(-0.5 \times 2\right) + \left(0.5 \times -2\right) = -2 \tag{10.66}$$

Equation (10.52) shows that CV_2 was −10°C and so will increase to zero as required. From Equation (10.55), CV_3 was originally zero and so will increase to 2 m³/hr – staying below the high limit of 3 m³/hr. Similarly CV_4 will reduce to −2 m³/hr – remaining above the low limit of −3 m³/hr.

If we were to reduce the maximum deviation from the average pass flow to 1 m³hr, then, assuming we give precedence to satisfying this constraint, ΔMV_1 would be limited to 1 m³/hr and ΔMV_2 to −1 m³/hr. Putting these values into Equation (10.59) gives

$$\Delta CV_1 = \left(-2.5 \times 1\right) + \left(2.5 \times -1\right) = -5 \tag{10.67}$$

In other words CV_1 would reduce from 10°C to 5°C and only half the temperature imbalance would be eliminated.

If CV_1 reaches zero, then so must CV_2. Similarly if CV_3 reaches the maximum deviation, then so must CV_4. We could omit CV_2 and CV_4 from the controller. Indeed, there is an argument for doing so; any inconsistency in the process gains or constraints could prevent the controller meeting its objective. For heaters with a larger number of passes it is common to include CVs for every pass.

MPC has the advantage that pass balancing can be included in the same controller as that for the rest of the process. For example, if the heater is occasionally a hydraulic constraint on capacity, we would include the pass flow controller outputs as additional CVs. If the only constraint on increasing feed rate were a single controller output, then pass temperature balancing would automatically be relaxed to allow all the pass valves to open to their limit and so increase total capacity.

11

Compressor Control

Compressors fall into one of two fundamental types – positive displacement and turbo-machines.

- Positive displacement machines can be either reciprocating or rotary. They both trap the gas in a space and then force it into a smaller volume and so increase its pressure. Reciprocating machines trap the gas in a cylinder. Rotary screw compressors have two helical screws that mesh together to trap the gas. Rotary vane compressors have radial blades; with the rotor mounted in an eccentric housing so that as it rotates, the blades slide into slots in the rotor so reducing the volume of the trapped gas.
- Turbo-machines impart velocity to the gas and its momentum carries it into a narrowing space and so its pressure increases. These machines can be either *axial* (in which the flow is parallel to the shaft) or *centrifugal* (sometimes described as *radial*, in which the flow is at right angles to the shaft). There are also *mixed flow* or *diagonal* compressors where the gas flow is at an angle to the rotor of between 0 and 90°. Multi-stage turbo-machines, with inter-cooling, are common.

11.1 Polytropic Head

Compressor performance is quoted in terms of *polytropic head*. This is the work done on the gas and its definition is developed from basic gas laws. Firstly *Boyle's Law* states that the volume (V) occupied by a gas is inversely proportional to its absolute pressure (P).

$$V \propto \frac{1}{P} \tag{11.1}$$

Charles' Law states that that volume is directly proportional to absolute temperature (T).

$$V \propto T \tag{11.2}$$

These laws are combined into the *Ideal Gas Law*

$$PV = RT \tag{11.3}$$

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

R is the *Universal Gas Constant*. It has a value of 8.314 kJ/kg-mole/K (1.9859 BTU/lb-mole/R). The law, in its normal form, is written on a molar basis. On a weight basis we include the molecular weight (MW) or sometimes the number of kg-moles per kg (n).

$$PV = \frac{RT}{MW} = nRT \tag{11.4}$$

Gases with polar molecules, because of the resulting attractive forces, occupy a volume less than that predicted by the Ideal Gas Law. Similarly gases with large molecules occupy a greater volume. These effects are more pronounced, at higher pressure and lower temperature, as the molecules become more closely packed – especially as the point of liquefaction is approached. To accommodate this non-ideal behaviour *compressibility (z)* is introduced into the equation.

$$PV = \frac{zRT}{MW} \tag{11.5}$$

Compressibility is determined experimentally for each gas. As an example, that for air at different temperatures and pressures, is included as Figure 11.1. For a wide range of operating conditions, especially well away from liquefaction, z is close to 1. Air, made of up non-polar, relatively small, molecules behaves much like an ideal gas.

The law governing *isentropic* compression can be written as

$$PV^{\gamma} = constant \tag{11.6}$$

The *adiabatic index* (γ) is defined as the ratio of the specific heat of the gas measured at constant pressure (c_p) to that measured at constant volume (c_v).

$$\gamma = \frac{c_p}{c_v} \tag{11.7}$$

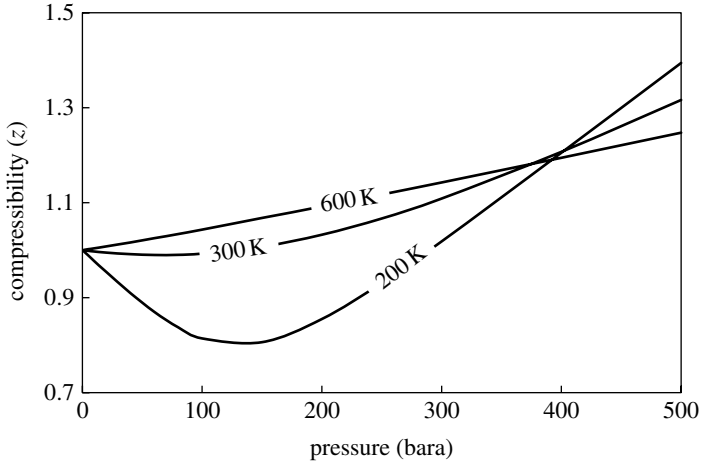

Figure 11.1 *Compressibility of air*

The value of γ is available for most gases in data books. For *monatomic* gases, i.e. those with one atom in their molecule such as He and Ar, it is typically 1.67. For *diatomic* gases, such as H_2, O_2, N_2 and CO, it is typically 1.40. For *triatomic* gases, such as H_2S and SO_2, it is around 1.33.

For compression to be described as isentropic (i.e. no change in entropy) firstly it has to be *reversible*. This means that the work done compressing the gas can be fully recovered by decompressing it. This is equivalent to the compressed gas entering the discharge of the compressor, being used to drive the machine backwards which then generates as much energy as was used to compress the gas. Due to losses in the machine, such as those caused by overcoming friction, reversibility is unachievable. Secondly the process has to be *adiabatic*, i.e. no heat must enter or leave the system. Again, due to heat lost to the atmosphere or to other compressor cooling systems, this condition will not be met. We describe a process where entropy changes as *polytropic*.

We can write Equations (11.3) and (11.6) for the suction (s) and discharge (d) of the compressor.

$$\frac{P_s V_s}{T_s} = \frac{P_d V_d}{T_d} \tag{11.8}$$

$$P_s V_s^{\gamma} = P_d V_d^{\gamma} \tag{11.9}$$

Combining these equations gives, for isentropic compression

$$\frac{P_d}{P_s} = \left(\frac{V_s}{V_d}\right)^{\gamma} = \left(\frac{T_d}{T_s}\right)^{\frac{\gamma}{\gamma-1}} \tag{11.10}$$

For polytropic compression we rewrite this equation using the *polytropic index* (n).

$$\frac{P_d}{P_s} = \left(\frac{V_s}{V_d}\right)^{n} = \left(\frac{T_d}{T_s}\right)^{\frac{n}{n-1}} \qquad n \neq \gamma \tag{11.11}$$

Polytropic efficiency (η_p) describes how close compression is to isentropic and is defined as

$$\eta_p = \frac{\left(\frac{n}{n-1}\right)}{\left(\frac{\gamma}{\gamma-1}\right)} \tag{11.12}$$

Polytropic head (H_p) is the work done on the gas; by definition this is given by

$$H_p = \int_s^d V.dP \tag{11.13}$$

Modifying Equation (11.9) for polytropic compression

$$PV^n = P_s V_s^n \qquad \text{or} \qquad V = P_s^{\frac{1}{n}} V_s P^{\frac{-1}{n}} \tag{11.14}$$

Substituting for V in Equation (11.13)

$$H_p = P_s^{\frac{1}{n}} V_s \int_s^d P^{\frac{-1}{n}} dP \tag{11.15}$$

Thus

$$H_p = P_s^{\frac{1}{n}} V_s \frac{n}{n-1}\left[P_d^{\frac{n-1}{n}} - P_s^{\frac{n-1}{n}}\right] = P_s V_s \frac{n}{n-1}\left[\left(\frac{P_d}{P_s}\right)^{\frac{n-1}{n}} - 1\right] \tag{11.16}$$

Substituting from Equations (11.12) and (11.5)

$$H_p = \eta_p \frac{\gamma}{\gamma - 1} \frac{z_s R T_s}{MW}\left[\left(\frac{P_d}{P_s}\right)^{\frac{\gamma-1}{\eta_p \gamma}} - 1\right] \tag{11.17}$$

Remember that pressure (P) and temperature (T) are on an absolute basis. Polytropic head (H_p) will have units such J/kg or BTU/lb.

Figure 11.2 shows a typical set of compressor performance curves. The performance curve for a centrifugal compressor, unlike that for a pump, terminates before the flow reaches zero. At too low a flow the compressor will *surge*. The impeller discharge pressure temporarily falls below that in the discharge pipework causing a transient flow reversal. This causes large and rapid fluctuations in flow and pressure. It can also be extremely noisy, although much of the noise may arise from the check (non-return) valve in the discharge pipework opening and closing rapidly. On some machines the resulting vibration can damage the compressor and/or its gearbox very quickly. Others tolerate the condition for longer.

The upper limit of the performance curve is *stonewall*. This arises when the speed of the gas, relative to the impeller, approaches the speed of sound (at conditions within the machine). Since gas cannot travel faster than this, the maximum capacity of the machine has been reached. No immediate damage to the machine is likely under these conditions, although over-heating may occur if operated at a high load for a long period.

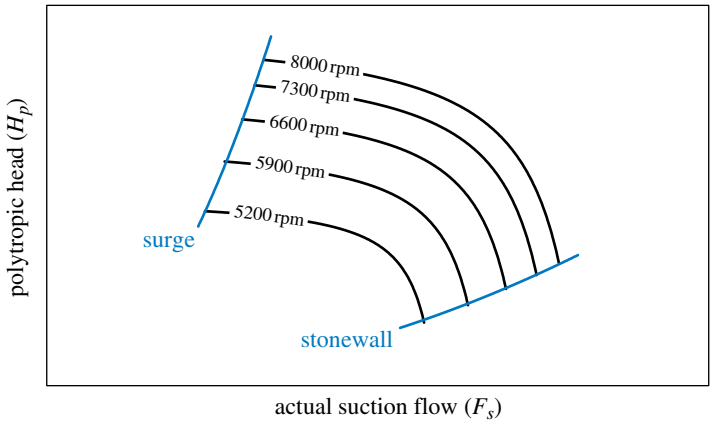

Figure 11.2 *Compressor performance*

11.2 Load Control (Turbo-Machines)

The term *load control* is a general term referring either to flow control or to pressure control. In some situations, such as compressing gas as a feed stream to a process, the requirement will be to control the flow of gas. In others, such as the compression of gas supplying a gas header, the requirement will be to control the compressor discharge pressure. Throughout this chapter it is assumed we require flow control. However, all of the schemes described can be modified to provide pressure control.

Compressor performance curves should strictly be plots of polytropic head (H_p) against suction flow measured in actual volumetric units (F_s). However it is common for discharge pressure, or the ratio of discharge to suction pressure, to replace polytropic head. In order to simplify the description of how possible flow controls operate, we will use the curves in this form. These approximations assume suction conditions and molecular weight remain constant. Should either change, then the performance curve, drawn on this basis, will move – as illustrated (for a fixed speed machine) in Figure 11.3.

Figure 11.4 shows a simplified process diagram showing a fixed speed compressor delivering gas to a downstream process. This pressure drop through the process increases in proportion with the square of the flow (measured in actual volumetric units). Assuming the pressure at the exit of the process is constant, then we can add the *process curve* to the compressor performance curve. The discharge pressure when the flow is zero is the *static head* and so the process curve will not usually pass through the origin. Since the flow through the compressor must equal the flow through the process, and the compressor discharge pressure must be the same as the process inlet pressure, the compressor will operate where the two lines cross. In order to adjust the flow we have to cause either the process curve or the compressor curve to move. Alternatively we have to remove the condition that discharge and process inlet pressures are equal, or that the flows are equal.

Figure 11.5 shows the first of several possible schemes. By placing a control valve between the compressor and the process the pressures are no longer equal. Or, if we think of the valve as now part of the process, the process curve will now move as we change the valve opening. Because of the nonlinear behaviour of both the compressor and the process the relationship between flow and valve Δp is highly nonlinear. An equal percentage valve or some other form of controller output conditioning (as described in Chapter 5) can be used to avoid problems in tuning the flow controller. As might be expected, because

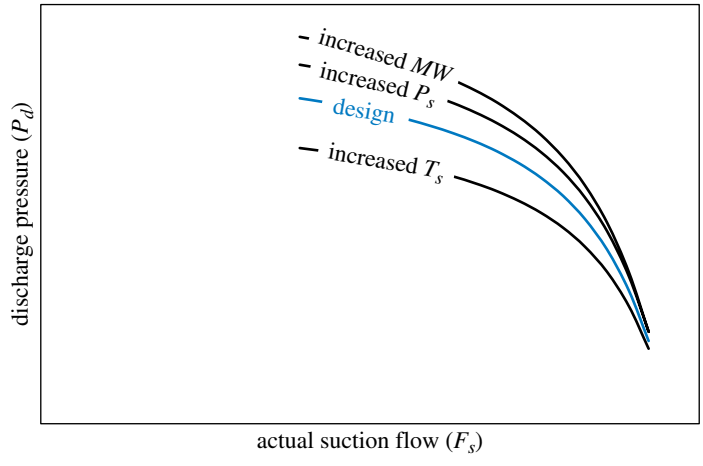

Figure 11.3 *Impact of suction conditions on compressor performance*

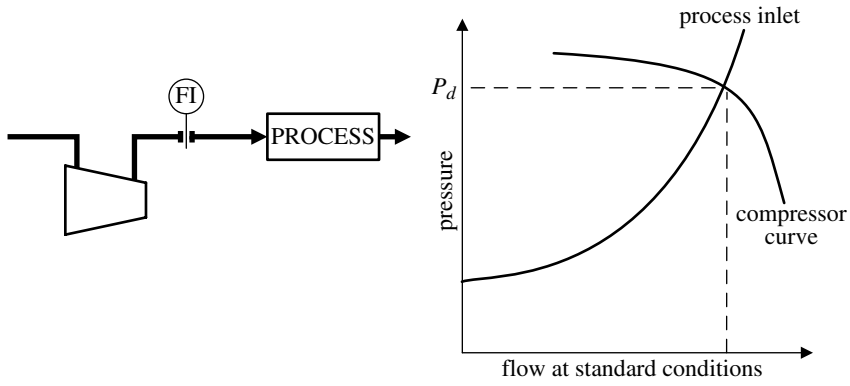

Figure 11.4 *Process diagram*

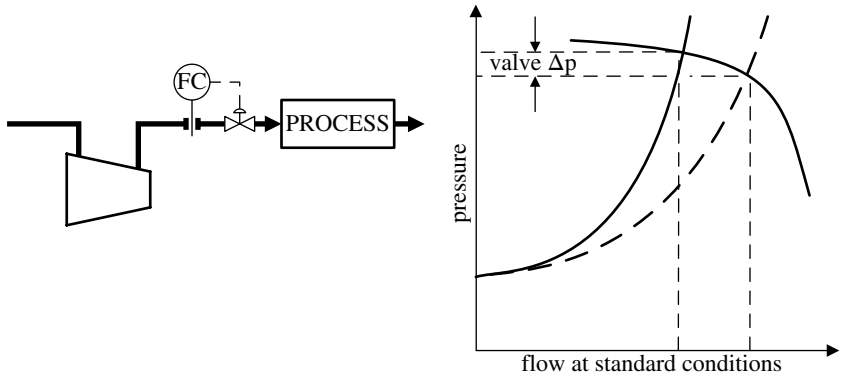

Figure 11.5 *Throttling discharge*

we are expending energy to raise the pressure of the gas only to partially reduce it again across the valve, the scheme is not energy efficient. Indeed, polytropic head rises rapidly as the discharge valve is closed because of the increase in discharge pressure. The increase in work done per unit mass of gas can outweigh the saving from the reduction in gas flow so that the net effect is an increase in power requirement. The range of the control is limited since, as we close the valve, the process curve approaches the end of the compressor curve and surge occurs.

Figure 11.6 shows how the control valve can be relocated to the suction of the compressor. Because we have plotted the compressor curve in terms of discharge pressure, rather than polytropic head, the curve moves when we change suction pressure (as previously described by Figure 11.3). Because of the lower pressures involved this is more energy efficient than discharge throttling. It also has a greater range because, while throttling reduces the normalised flow, the actual suction flow increases as suction pressure is reduced. This helps keep the machine away from surge. One concern is that it is possible for the suction pressure to fall below atmospheric pressure; any leaks in the pipework would then permit air to enter the compressor. If compressing a flammable gas this could cause an explosion inside the machine and so some form of minimum suction pressure override would be necessary. It is also common to install oxygen detection on such machines.

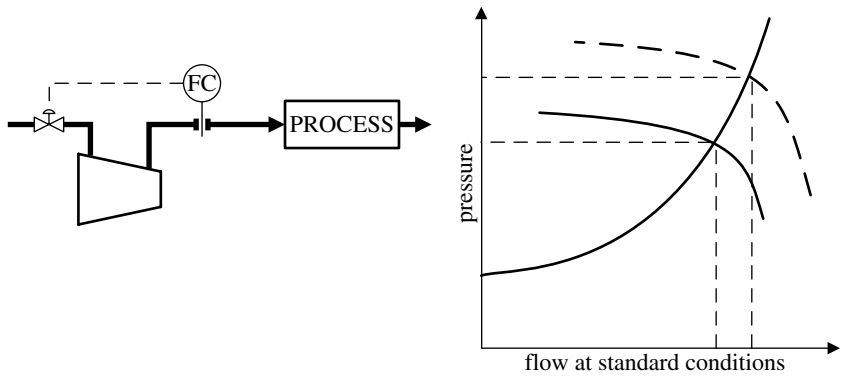

Figure 11.6 Throttling suction

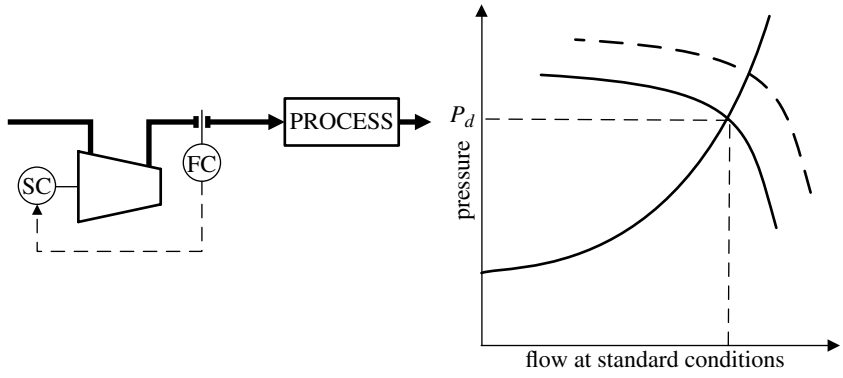

Figure 11.7 Adjusting speed

Changing speed is another way of moving the compressor curve. We can use this to control the flow, as shown in Figure 11.7. This is energy efficient and, because the surge point moves, can operate over a wide range. However, variable frequency drives (VFD) for large electric motors are costly. Variable speed steam turbine drivers have a mixed reputation. Most success has been had with gas turbine drivers.

Figure 11.8 shows the use of inlet guide-vanes (IGV) on a constant speed machine. These convert the inlet gas's forward momentum into rotational momentum. The angle of the guide-vanes is adjustable; conventionally negative angles give *pre-rotation* and positive angles produce *counter-rotation*. Pre-rotation increases the compressor efficiency permitting it to deliver a greater flow at the same discharge pressure. They are therefore also an energy efficient method for flow control. But, at steep angles, the guide-vanes effectively behave like suction throttling. Because the guide-vanes are inside the compressor, adjusting them changes the geometry of the machine itself. For variable speed machines, a family of compressor curves will exist for each guide-vane angle – each with a different surge point. Mechanically guide-vanes are more complex and potentially more costly to maintain.

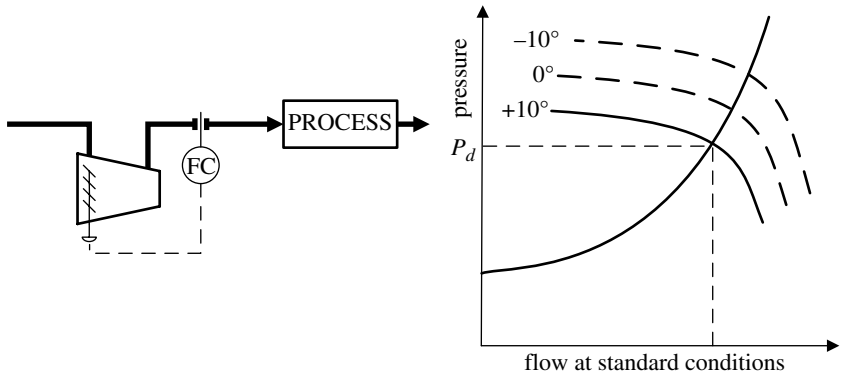

Figure 11.8 *Inlet guide vanes*

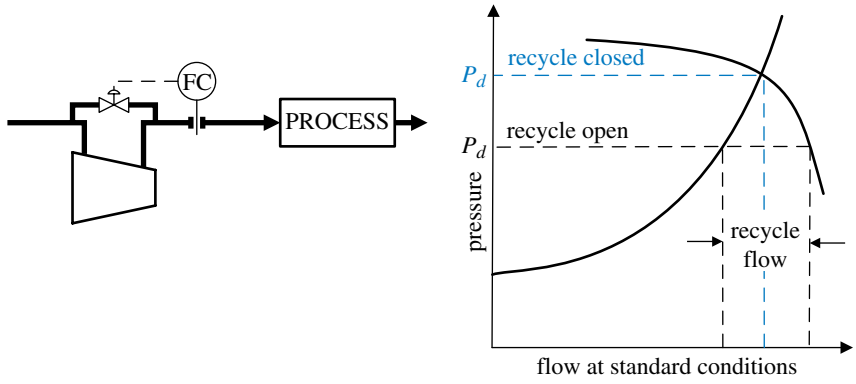

Figure 11.9 *Recycle manipulation*

By partially recycling gas through the compressor we remove the condition that the flow through the machine must equal the flow through the process. The scheme, shown in Figure 11.9, offers the greatest range since the surge point is never approached. It is, however, costly to install and operate. In reality the system is more complex than drawn; for example, the recycle must be cooled. It is the least energy efficient approach. It can, however, be used to increase the range of the other schemes described by recycling just enough to avoid surge.

Figure 11.10 shows the energy efficiency and operating limit of each of the schemes. Throttling the discharge is by far the least efficient with power consumption initially increasing as flow through the machine is reduced. Its range is limited by surge. Manipulating recycle overcomes this limit but is, as expected, very inefficient. Throttling the suction is far more energy efficient and permits operation at lower loads than discharge throttling. The use of inlet guide-vanes and manipulation of speed are the most efficient. Reducing speed reduces the flow at which surge occurs and so also increases the operating range, although there may be minimum and maximum speed constraints that prevent this from being fully exploited. The operating range of the machine can be extended by increasing suction pressure.

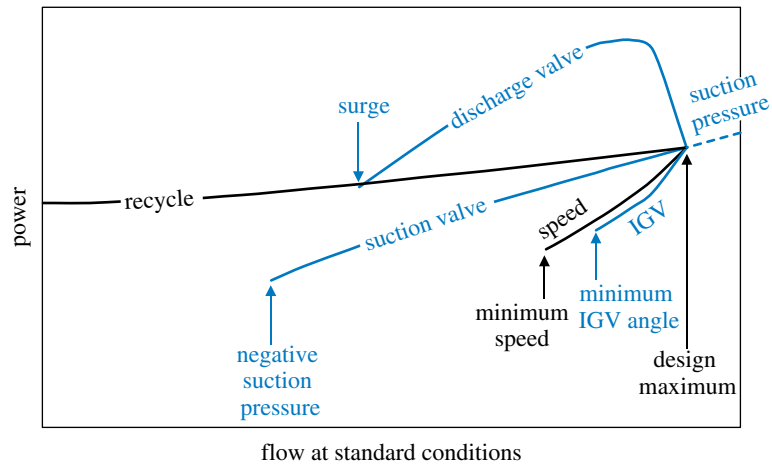

Figure 11.10 *Energy efficiency*

11.3 Load Control (Reciprocating Machines)

Many of the schemes described for turbo-machines can be applied to reciprocating machines. Suction throttling reduces the suction pressure and so less gas (on a weight basis) enters the cylinder and thus less gas will be discharged. But discharge throttling has little effect on flow. Once in the cylinder the machine will deliver a fixed volume of gas at each stroke. Normally the volume of gas remaining in the cylinder at the end of the compression stroke will be close to zero. Throttling simply means that the machine has to work harder to overcome the restriction. However, if there is space left in the cylinder at the end of the compression stroke, at a higher discharge pressure a greater mass of gas will remain in the machine and so the flow delivered will be reduced.

Adjusting speed is an energy efficient means of controlling flow. Recycling is effective but, like turbo-machines, is costly.

A further option is *cylinder loading* which alters the effective compression ratio. Reciprocating machines follow the cycle shown in Figure 11.11. At point A the piston is at *top dead centre (TDC)*, having just finished discharging gas. The machine then begins the suction stoke. At point B the inlet valve opens and then closes at point C when the piston reaches *bottom dead centre (BDC)*. It then begins the compression stroke with the exhaust valve opening at D and then closing at A.

Cylinder loading can take one of two forms. *Valve loading* changes the point in the cycle at which the inlet valve closes. If set at 75% load, on reaching BDC the valve remains open. The piston begins the compression stroke but no actual compression takes place. Instead gas leaves via the inlet valve until it closes at point C_{75}. At this point true compression begins but on a gas volume that is 75% of maximum. The load can similarly be reduced to 50% or 25% of maximum. The alternative approach is the use of a *head end unloader*. This comprises an adjustable piston fitted inside the cylinder head. Moving it changes the volume of gas in the cylinder at TDC. Typically it allows adjustment of flow down to about 75% of maximum – and so provides much less flexibility than valve loading. It does, however, permit much finer adjustment of flow. Cylinder loading is normally set manually local to the machine. However, there are examples of it being automated to provide true flow control.

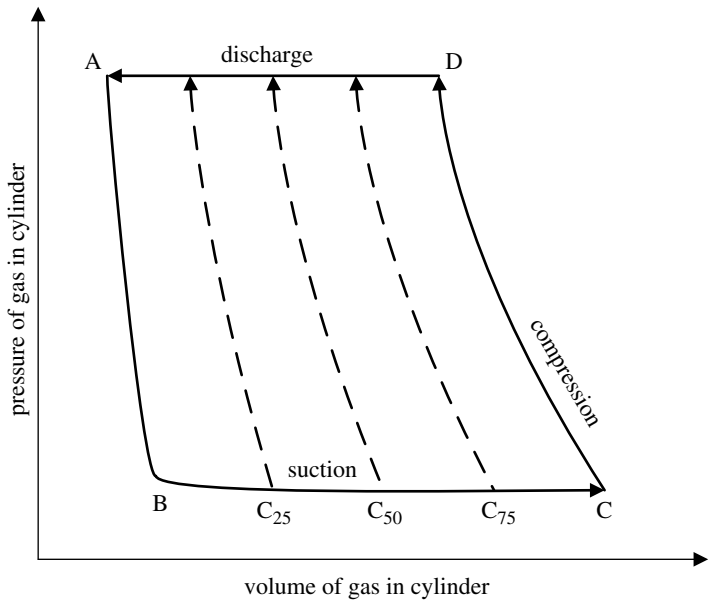

Figure 11.11 *Cylinder loading*

11.4 Anti-Surge Control

Surge is not a problem with positive displacement machines. On turbo-machines it can be avoided by recycling but this is costly. However, the cost of repair and lost production that can arise from surging a compressor may be greater. The objective is to minimise recycling without jeopardising the machine. To achieve this we need to be able to predict, using the available instrumentation, that surge is about to occur.

The complexity of the anti-surge scheme depends much on the machine and its duty. For example, a machine running at a fixed speed and compressing a gas of constant composition might require only minimum flow protection, as shown in Figure 11.12.

The flow measurement is normally located on the suction side, since compressor curves are normally presented in terms of actual suction flow. But, providing suction and discharge pressures are constant, the scheme will operate equally effectively if the flow measurement is downstream of the compressor. However it relies on there being a significant change in flow as the surge point is approached, i.e. the compressor curve is reasonably 'flat' in this area. This is generally the case for centrifugal machines. For axial machines, which have a steeper curve, it is likely that the scheme shown in Figure 11.13 is preferable. This will cause the recycle to open if the discharge pressure exceeds SP. Here pressure must change significantly as surge is approached and so the scheme is better for 'steep' compressor curves.

For variable speed machines surge is a line, not a single point. The next group of schemes are based on developing an equation for this line – either from the compressor manufacturer's data or from plant testing. The surge line can usually be represented by a quadratic function.

$$H_p = f\left(F_s^2\right) \tag{11.18}$$

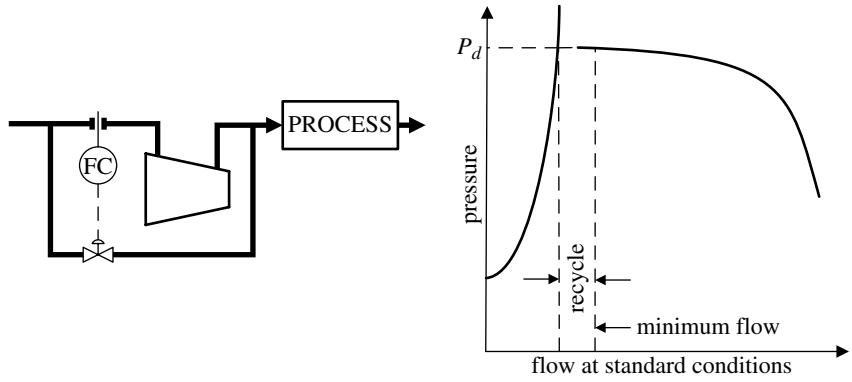

Figure 11.12 *Minimum flow protection*

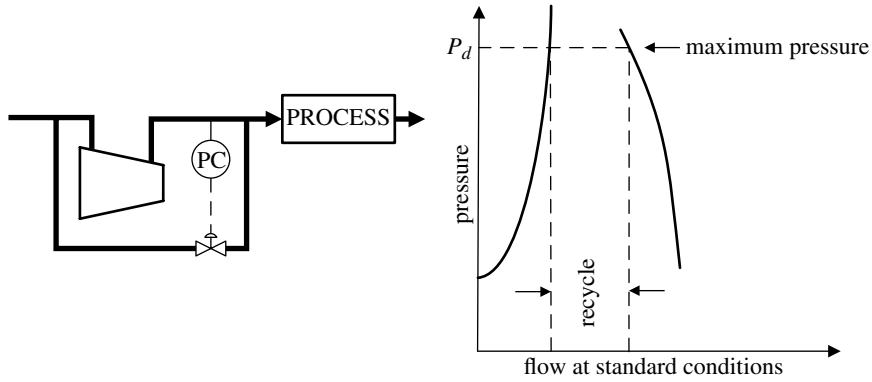

Figure 11.13 *Maximum pressure protection*

Figure 11.14 shows the compressor curves now redrawn by plotting against F_s^2 instead of F_s. The surge line is now approximately straight. A second (dashed) line has been added, building in a safety margin of 15% over the minimum flow. F_s is measured by an orifice plate type meter and is related to the pressure drop (dp) across the orifice.

$$F_s = c_d \frac{\pi d^2}{4} \sqrt{\frac{2dp}{\rho_s}} \tag{11.19}$$

If we assume the discharge coefficient (c_d), the pipe diameter (d) and the fluid density (ρ_s) are all constant then building in the 15% margin gives

$$\left(1.15F_s\right)^2 \propto 1.15^2 dp \tag{11.20}$$

Or, if m is slope of the minimum flow line and c the intercept on the polytropic head axis, then

$$H_p = m.\left(1.15^2 dp\right) + c \tag{11.21}$$

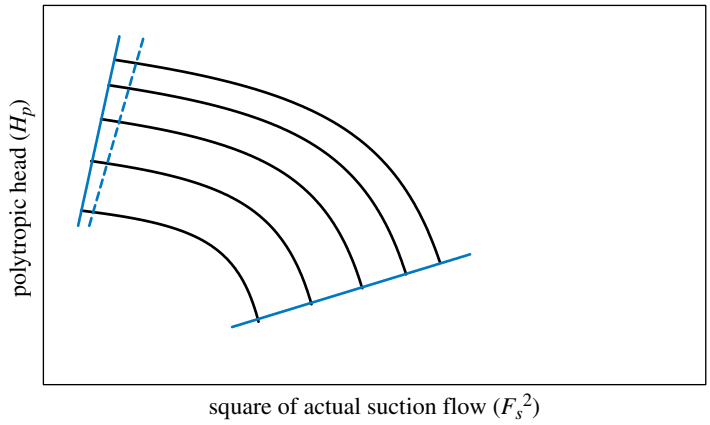

Figure 11.14 *Linearised surge line*

To apply this technique we need to be able to measure H_p and dp. If we have a suction flow meter then the measurement of dp is readily available. We need only to compensate for any square root extraction that may be in the control system (see Chapter 5). However the measurement of H_p is not so straightforward. Equation (11.17) includes values which we cannot measure – such as polytropic efficiency. While we can assume values for z and γ, in fact they can change with pressure, temperature and composition. We can, of course, measure MW (as we did in Chapter 10) but less reliably than other process measurements.

There are a number of possible approaches which use a parameter related to H_p. A common example is the pressure rise $(P_d - P_s)$. By plotting $(P_d - P_s)$ against $1.15^2 dp$ we can determine m and c but we need a number of points collected close to surge. It may be possible to develop a correlation from information provided by the compressor manufacturer. Or, if the machine has surged on previous occasions, the process data may exist in the plant history database. Failing either of these approaches it would be necessary to obtain data by testing close to the surge point – a process that needs to be managed carefully.

Historically this type of scheme has been implemented as shown in Figure 11.15. Unusually the controller SP is a process measurement – in this case $(P_d - P_s)$. The controller PV is the value that $(P_d - P_s)$ would have if the compressor were operating at the minimum flow. The controller, by manipulating the recycle, will attempt to eliminate the difference between the PV and the SP. If recycle is needed to avoid surge then, by doing so, this will ensure that the compressor operates at the minimum flow. If recycle is not needed then the controller will fully close the recycle valve leaving an offset between the PV and the SP.

While effective, the rather unconventional approach requires a slightly different approach to implementation. Firstly, when obtaining the process dynamics for tuning purposes by stepping the recycle valve, both the SP and the PV will change. Model identification therefore should be based on the difference PV – SP (the error). Secondly, the controller cannot initialise using PV tracking. If switched to automatic when the compressor is operating above minimum flow, with the recycle valve closed, the controller will do nothing. If the flow is below minimum, then the controller will immediately take corrective action – as required. However, if the recycle is unnecessarily open, then the corrective action taken by the controller can cause a process upset. Finally, the unconventional approach is a disadvantage

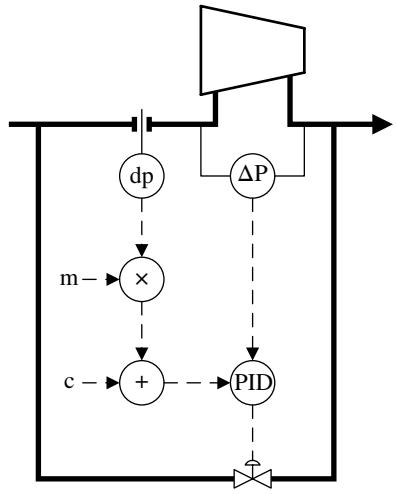

Figure 11.15 *Implementation of anti-surge control*

for operator understanding and does not permit any adjustment by the operator to how closely surge is approached.

An alternative approach is to rearrange the equation of the minimum flow line.

$$\left(P_d - P_s\right) = m.dp + c \tag{11.22}$$

Substituting for *dp* in Equation (11.19)

$$F_{min} = c_d \frac{\pi d^2}{4} \sqrt{\frac{2\left(\frac{P_d - P_s - c}{m}\right)}{\rho}} \tag{11.23}$$

This is then a measure of the minimum suction flow. This value could be used as the SP of a minimum flow type anti-surge controller. Or it could be subtracted from the measured flow to give the margin to surge in flow units.

$$F_{margin} = F_s - F_{min} \tag{11.24}$$

This would then be the PV of the anti-surge controller with the operator entering the required margin as the SP.

It should be emphasised that the use of the pressure rise across the machine is just one possible approximation to polytropic head. Pressure rise is a special case of the function ($aP_s + bP_d + c$), where the coefficients *a*, *b* and *c* are chosen for each machine. Another approach is to use the pressure ratio (P_d/P_s) – remembering that, unlike the pressure rise, it is important here to convert to absolute pressures. Other schemes use the speed of the machine or its power consumption. Further, any of these approximations to H_p can be plotted against some nonlinear function of *dp* rather than just its

measurement. The most effective solution is machine-specific and each should be evaluated before building the controller.

There is however a more rigorous approach. By combining Equations (11.11) and (11.12) we get

$$\left(\frac{P_d}{P_s}\right)^{\frac{\gamma-1}{\eta_p \gamma}} = \frac{T_d}{T_s} \tag{11.25}$$

Rearranging we get

$$\eta_p \frac{\gamma}{\gamma-1} = \frac{\log\left(\frac{P_d}{P_s}\right)}{\log\left(\frac{T_d}{T_s}\right)} \tag{11.26}$$

Substituting Equations (11.25) and (11.26) into Equation (11.17) we get an alternative definition of polytropic head

$$H_p = \frac{\log\left(\frac{P_d}{P_s}\right)}{\log\left(\frac{T_d}{T_s}\right)} \frac{z_s R T_s}{MW}\left[\frac{T_d}{T_s} - 1\right] \tag{11.27}$$

From Equation (11.19)

$$F_s^2 = 2\left(\frac{c_d \pi d^2}{4}\right)^2 \frac{dp}{\rho_s} \tag{11.28}$$

The gas density (ρ_s) is inversely proportional to specific volume (V). So, from Equation (11.5), we get

$$\rho_s = \frac{MW.P_s}{z_s T_s} \tag{11.29}$$

Replacing ρ_s in Equation (11.28) gives

$$F_s^2 = 2\left(\frac{c_d \pi d^2}{4}\right)^2 \frac{dp.z_s T_s}{MW.P_s} \tag{11.30}$$

We assume that the surge line can be represented by a quadratic, i.e. H_p varies linearly with F_s^2 and so, since H_p is zero when F_s^2 is zero, the slope of the surge line is given by

$$\frac{H_p}{F_s^2} = \frac{R}{2}\left(\frac{4}{c_d \pi d^2}\right)^2 \frac{\log\left(\frac{P_d}{P_s}\right)}{\log\left(\frac{T_d}{T_s}\right)} \frac{P_s}{dp}\left[\frac{T_d}{T_s} - 1\right] \tag{11.31}$$

We do not need to know the absolute value of the slope, only that it is changing, so terms which are constant can be omitted.

$$\frac{H_p}{F_s^2} \propto \frac{\log\left(\frac{P_d}{P_s}\right)}{\log\left(\frac{T_d}{T_s}\right)} \frac{P_s}{dp}\left[\frac{T_d}{T_s} - 1\right] \tag{11.32}$$

Terms that are difficult to measure, such as polytropic efficiency (η_p), compressibility (z_s), adiabatic index (γ) and molecular weight (MW), have now been eliminated. Provided there are measurements of suction flow; suction and discharge pressure; and suction and discharge temperature, we can calculate a parameter which is proportional to the slope of the surge line. This parameter is used as the PV of the anti-surge controller. The value of this parameter at which the compressor surges can be predicted by using data collected historically or by taking the machine into surge. Remembering to move the machine away from surge the slope of this line is reduced. Thus, to incorporate a safety margin, the SP of the controller is set to 15% (say) less than the slope at which surge is known to occur.

This scheme should prove more robust as gas composition and other operating conditions vary. It is thus often used in proprietary compressor control systems.

It is important that the controllers in all the anti-surge schemes described have anti-reset windup. Unless the required gas flow is always below the surge point the controller will mostly be operating with the recycle valve fully shut, i.e. it will be saturated. The valve needs to open quickly when surge is approached and any windup could result in the compressor surging before the valve opens. For the same reason it is also common for quick opening valves (as described in Chapter 5) to be used. These present a tuning challenge; the nonlinearity they introduce means that the process gain is highest when they initially open. The anti-surge controller needs to be tuned therefore for operation close to this condition; otherwise it may become unstable. Obtaining the process dynamics from plant testing also needs to be approached with care since the tests should ideally be performed with the machine close to surge.

In addition to anti-surge control, it is common to have a surge recovery scheme. This relies on the detection of rapid changes in pressure or flow and then overrides the anti-surge scheme to fully open the recycle valve. It can also increase the safety margin used by the anti-surge controller so that, when the anti-surge scheme is permitted to take back control, the same situation should not recur.

The anti-surge controller will interact with other schemes on the compressor. For example, it opening the recycle valve will cause the discharge pressure to drop and hence also the flow to the downstream process. The flow (or pressure) controller will then take corrective action. To avoid this interaction causing an oscillatory response the load controller should be tuned to act significantly slower than the anti-surge scheme.

In multi-stage machines, the anti-surge scheme on the first stage will temporarily starve the next stage(s) of gas – possibly causing them to approach surge. Feedforward compensation may be required to avoid significant disturbances.

Special consideration must be given to compressors operating in parallel. Even so-called identical machines will have slightly different compressor curves. The machines must operate with the same inlet pressures and the same outlet pressures. If operating on a flat section of the curves, equalising the pressure rise may result in very different flows through each machine. For example, Figure 11.16 shows the effect of a small change in the molecular weight of the gas. This causes the compressor curves to move vertically. As a result the flow though machine A moves from being very close to surge to being very close to stonewall.

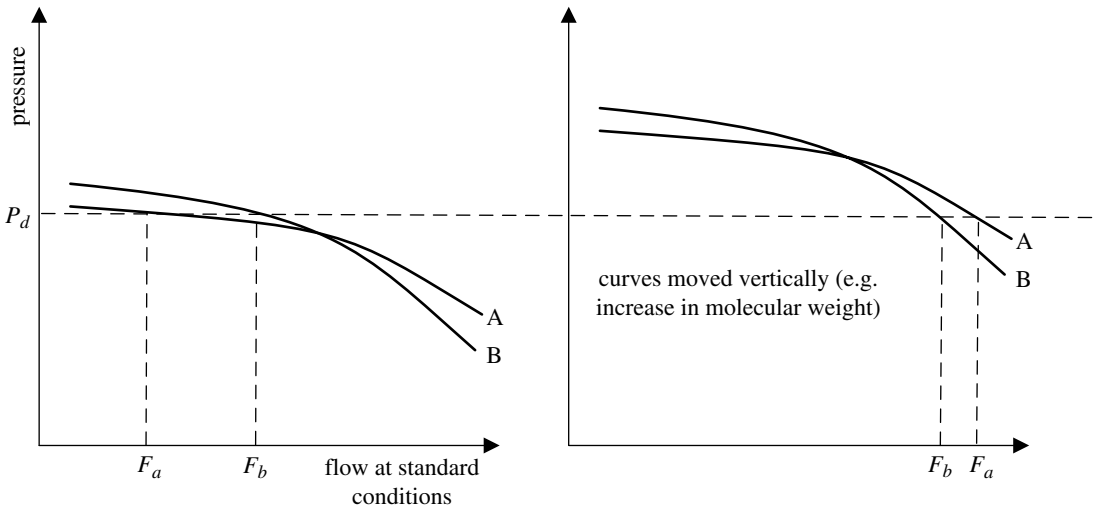

Figure 11.16 Problem of parallel compressors

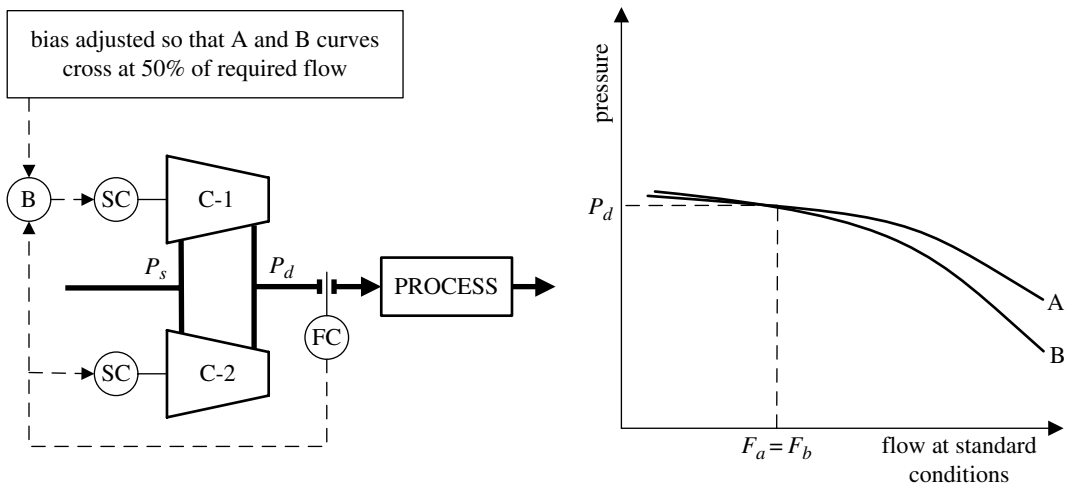

Figure 11.17 Balancing parallel compressors

It is also possible that the load will swing between the machines – particularly during process disturbances or compressor start-up. It is important that the flow control strategy balances the machines – for example, by applying a bias to one of the speed controllers, so that both machines operate at roughly the same distance from surge. Figure 11.17 shows a possible scheme.

12

Distillation Control

Figure 12.1 shows an example of the internal arrangement of a common type of distillation column. Liquid from the *downcomer* flows across the *tray*. Vapour from the tray below passes through the liquid – exchanging less volatile components for more volatile ones from the liquid. Trays are numbered but there is no convention as whether they are counted from the top or the bottom of the column. In this book we number them from the bottom.

When counting trays in a column we generally refer to actual trays. We can also count *theoretical trays* (also known as *theoretical stages*). Actual trays are not 100% efficient; theoretical trays are. We can convert from actual to theoretical trays by multiplying by the tray efficiency. There are columns where actual trays are replaced by packing; then we would refer to theoretical trays per unit height of packing.

Although not used in this book, other texts refer to the section of the column above the feed tray as the *rectifying* or the *enriching* section. In this section the vapour flow is greater than the liquid flow. The lower section is referred to as the *stripping* or *exhausting* section. Here the vapour flow is smaller than the liquid flow.

A typical external arrangement is shown in Figure 12.2. Feed enters the column on the *feed tray*. Depending on its enthalpy it will split between vapour and liquid. Vapour joins that rising from the *reboiler* and travels up the column to the condenser. A *total condenser*, as the name suggests, condenses all the vapour. A *partial condenser* does not, resulting in the need to withdraw a gaseous product from the *reflux drum*. This drum, also known as the *reflux accumulator* or *overheads drum* collects the condensed liquid. Part is pumped back to the column as *reflux*, the remainder leaves as *distillate* – also known as *overheads*. Whether this needs to be pumped will depend on the downstream pressure. The liquid portion of the feed joins the reflux and leaves the base of the column. Part of it is evaporated by the *reboiler*; the remainder leaves as *bottoms*. Again the need for pumping will depend on the downstream pressure.

Figure 12.3 describes the column's operating envelope. *Blowing* can occur when the internal vapour rate is high. Instead of the vapour breaking up into bubbles to produce foam on each tray, its velocity is such that it produces 'tubes' of vapour through the liquid on the tray. The surface area of vapour/liquid contact is substantially reduced and the tray efficiency falls.

Conversely, *weeping* can occur when the vapour rate is very low. Without sufficient vapour passing through the tray, liquid instead weeps through the tray. It effectively bypasses the tray and again the tray efficiency falls.

Blowing and weeping are unusual in conventional distillation columns. To produce more vapour we increase the reboiler duty but, in order to maintain the heat balance, we increase the condenser duty.

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

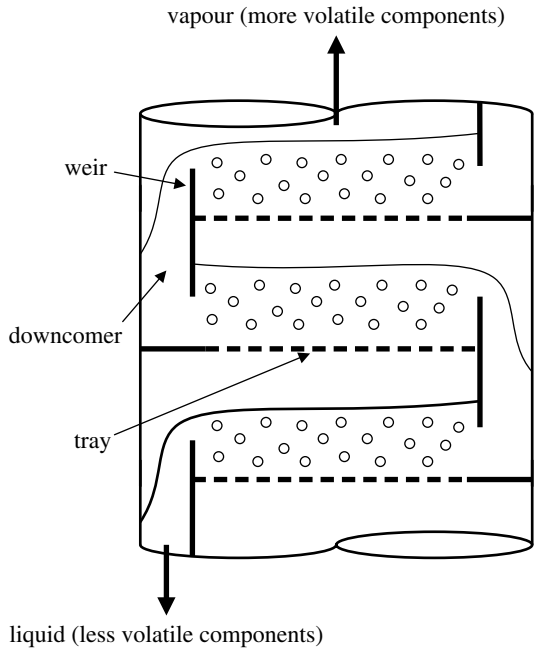

Figure 12.1 *Column internals*

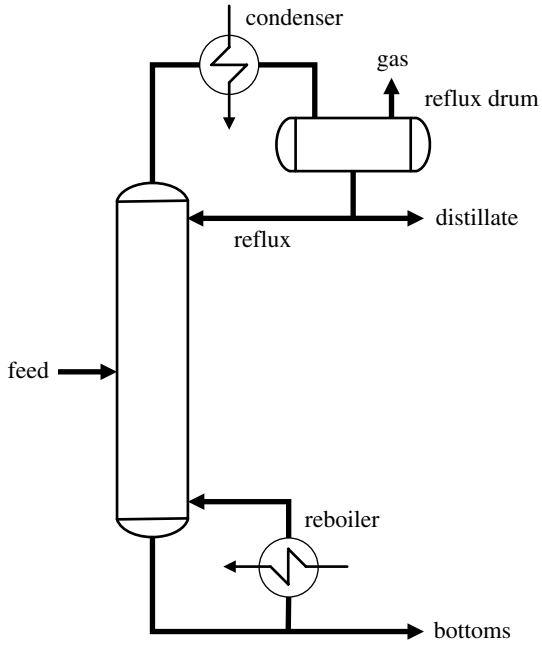

Figure 12.2 *Basic column*

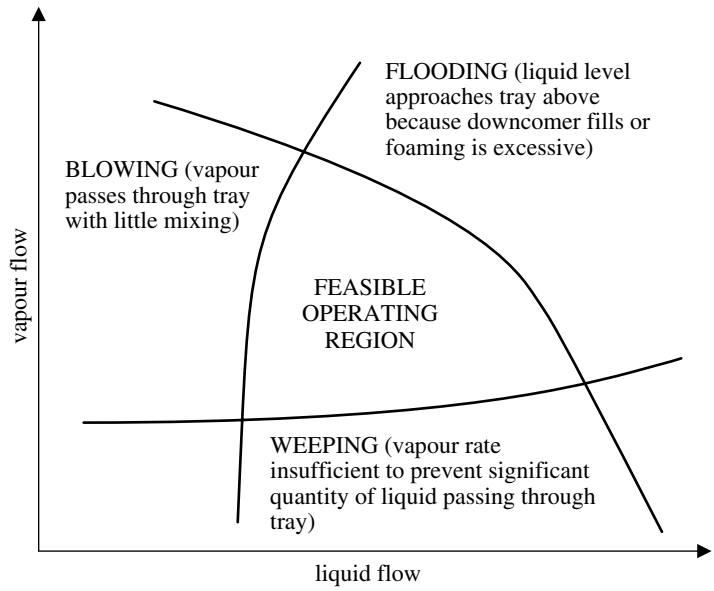

Figure 12.3 *Operating envelop*

Depending on the configuration of the controls, the additional liquid formed in the reflux drum will be returned to the column. Thus vapour and liquid rates change almost in proportion – avoiding the blowing and weeping constraints. However, there are columns where the two flows are not related. For example, reflux may be provided from some external source – as it is in gas scrubbing columns.

Flooding is the most common of the hydraulic constraints likely to be encountered. There are two mechanisms that cause it. The first, *downcomer flooding*, arises if the maximum internal liquid rate is exceeded. Liquid flows through the downcomer under gravity. The level of liquid built up in the downcomer is the result of a balance between the pressure drop across it and the head of liquid held above it on the tray. As the flow increases, the pressure drop increases (with the square of the flow) and so the head must increase. Ultimately the level reaches the tray above and the tray ceases to provide any separation.

Jet flooding is caused by excessive vapour rates or foaming. Increasing the vapour mixed with the liquid on the tray decreases the density of the foam. Thus a greater depth is required to provide the head necessary for the liquid to pass down the downcomer. Again the level of the foam will ultimately reach the tray above.

The problem is exacerbated by the composition controllers on the column. The drop in tray efficiency will result in a reduction in product purity. The controllers will respond by increasing reboiler duty and/or reflux, thus further increasing the internal traffic. While flooding may have occurred in only a small section of the column, without intervention it will quickly propagate to the whole column.

From a control perspective, accurate detection of flooding is difficult. One approach is to monitor the pressure drop across the column. However, this is usually too late an indication; flooding is well established by the time a significant change is detected. A better approach is to measure the pressure drop across small sections of the column known to flood first. However, this too can be unreliable; it is common for the column to be operating normally with a high pressure drop one day, and to be flooding at a lower pressure drop on another. Attempts have been made to use tray loading calculations, usually used for column design, as a means of quantifying approach to flooding but these have proved unreliable. Using any of these techniques as an override to restrict the column will over-constrain the operation.

12.1 Key Components

Most distillation processes are multi-component. A column common to many industries is the *LPG splitter*. The feed composition is shown in Table 12.1. It is primarily a propane/butane mixture which is required to be separated to specified purity targets. While at first glance the feed appears to comprise four components, butane has two isomers (isobutane and n-butane) and pentane has three. Further it is likely that small amounts of unsaturated material will be present, such as C_3H_6, C_4H_8, C_5H_{10} and their isomers. Thus an apparently binary distillation actually involves a substantial number of components.

While the theory clearly exists to allow such columns to be designed, it is complex and its use here will make difficult the explanation of how the control strategies operate. We will therefore simplify the approach by treating the column as a binary separation. We do this by identifying the *key components*. The *light key (LK)* is the lightest component that will be found in any quantity in the bottom product. The *heavy key (HK)* is the heaviest that will be found in the distillate product.

As an example let us assume that the distillate must be 95% pure propane and the bottoms butane must contain 10% propane. Table 12.1 shows the resulting products. While the lightest in the bottoms is ethane, it is present only in a small concentration. The main impurity is propane. Similarly butane is the main impurity in the distillate. We therefore define these as the light and heavy key components. To close the mass balance we lump together all the components lighter than the light key component as the *light light key (LLK)* component and all those heavier than the heavy key component as the *heavy heavy key (HHK)* component.

12.2 Relative Volatility

The underlying principle of separating components by distillation is that, when a liquid is partially evaporated, the composition of the vapour produced is different from that of the liquid. One component must be more volatile than the other(s). Ease of separation depends on *relative volatility*.

There are a number of correlations which predict how pure components behave. The most commonly documented is the *Antoine Equation* which predicts the *vapour pressure* of the pure component (P_0) when at the temperature (T).

$$\ln\left(P_0\right) = A - \frac{B}{T + C} \tag{12.1}$$

A, B and C are constants determined experimentally. They are readily available from data books and the internet. They have engineering units so their numerical value will depend on the units of measure of

Table 12.1 *LPG splitter feed and product analysis*

Component	Feed	Distillate		Bottoms	
	moles/hr	moles/hr	mol %	moles/hr	mol %
C_2H_6	0.7	0.5	1.3	0.2	0.4
C_3H_8	45.0	39.1	95.0	5.9	10.0
C_4H_{10}	52.7	1.4	3.5	51.3	87.1
C_5H_{12}	1.6	0.1	0.2	1.5	2.5
Total	100.0	41.1	100.0	58.9	100.0

pressure and temperature. They also change if the Antoine Equation is based on the logarithm to base 10, i.e. $\log_{10}(P_0)$. Table 12.2 gives values for some common components. These assume pressure is measured in bara and temperature in °C.

As an aside, the Antoine Equation has a number of uses. Normal boiling point is defined as the temperature (T_b) at which the vapour pressure reaches atmospheric pressure (1.01325 bara). Rearranging Equation (12.1)

$$T_b = \frac{B}{A - 0.013163} - C \tag{12.2}$$

Of course, the boiling point can be determined at any pressure – a technique we shall use later to validate pressure compensation of tray temperatures.

Antoine's Law can also be used to determine *Reid Vapour Pressure (RVP)*. RVP is defined as the vapour pressure at 100°F (37.8°C).

$$RVP = e^{A - \frac{B}{37.8 + C}} \tag{12.3}$$

Table 12.3 shows the values of the properties predicted.

Returning to the use of Antoine in predicting volatilities, the coefficients in Table 12.2 were used to plot the curves in Figure 12.4. As might be expected, propene and propane have very similar vapour pressures and so would be difficult to separate. Propane and butane, however, would be a relatively easy separation.

To quantify this we will develop a definition of relative volatility from some basic equations of state. Firstly, *Raoult's Law* states that the partial pressure of a component *i* in the vapour (p_i) is proportional to its molar fraction in the liquid (x_i).

$$p_i = \left(P_0\right)_i x_i \tag{12.4}$$

For Raoult's Law to apply, the components need to chemically similar. As an improvement, for non-ideal systems, *Henry's Law* can be used. It uses the experimentally determined *Henry's coefficient (H)* as the constant of proportionality rather than the vapour pressure (P_0).

$$p_i = H_i x_i \tag{12.5}$$

Table 12.2 *Antoine coefficients*

Component		A	B	C
propene	C_3H_6	9.08250	1807.529	247.00
propane	C_3H_8	9.04654	1850.841	246.99
n-butane	C_4H_{10}	9.05800	2154.897	238.73

Table 12.3 *Predicted properties*

Component		T_b °C	*RVP* bara
propene	C_3H_6	−47.7	15.4
propane	C_3H_8	−42.1	12.8
butane	C_4H_{10}	−0.5	3.5

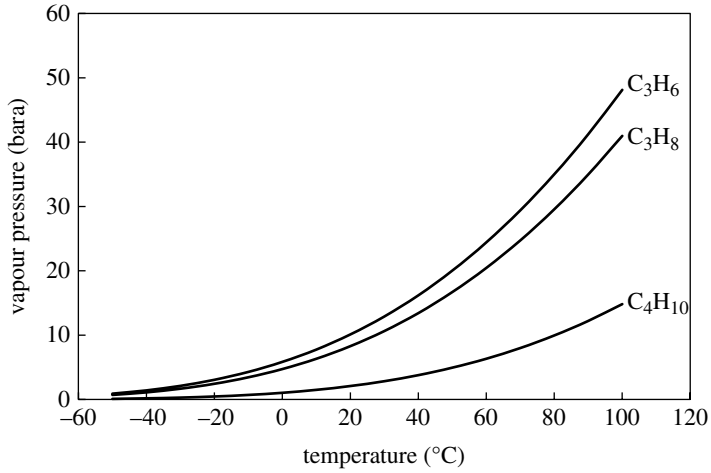

Figure 12.4 *Vapour pressure of common components*

Dalton's Law of Partial Pressures states that the partial pressures of the components sum to the total pressure (P).

$$\sum_{i=1}^{n} p_i = P \tag{12.6}$$

The *Ideal Gas Law* implies that partial pressure is proportional to the mole fraction (y_i) of the i^{th} component in the vapour.

$$p_i = Py_i \tag{12.7}$$

Combining these laws gives

$$y_i = \frac{(P_0)_i}{P} x_i \qquad \text{or} \qquad y_i = \frac{H_i}{P} x_i \tag{12.8}$$

Volatility (α) is defined as the ratio of the mole fraction of the component in the vapour to that in the liquid.

$$\alpha_i = \frac{y_i}{x_i} = \frac{(P_0)_i}{P} \qquad \text{or} \qquad \frac{H_i}{P} \tag{12.9}$$

Relative volatility is the ratio of the volatility of one component to that of another.

$$\alpha_{ij} = \frac{\alpha_i}{\alpha_j} = \frac{(P_0)_i}{(P_0)_j} \qquad \text{or} \qquad \frac{H_i}{H_j} \tag{12.10}$$

However, the required Henry's coefficient may not be published and, even if available, its variation with temperature may not be known. Since we are generally working with components that are chemically similar

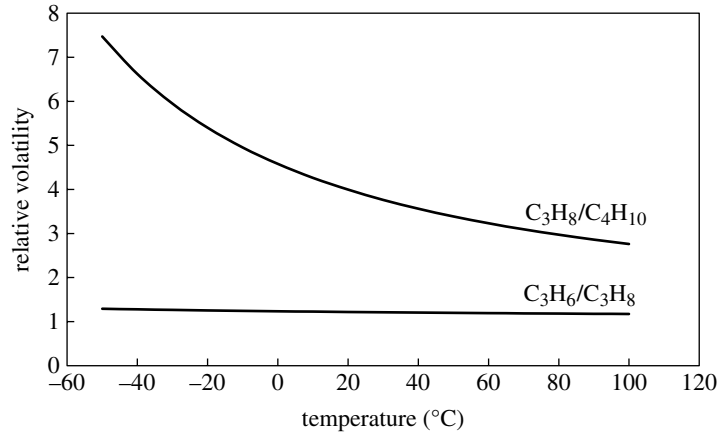

Figure 12.5 *Relative volatility of common mixtures*

we instead assume ideal behaviour and use the first definition of α_{ij}. Figure 12.5 shows how relative volatility varies with temperature. A value of 1 for relative volatility indicates that the components cannot be separated by distillation because the composition of the vapour evaporated is identical to that of the liquid. As expected the relative volatility for propene/propane is close to this value. That for propane/butane is far greater.

We will address later, in the section covering optimisation, the variation of relative volatility with temperature. Since liquids in the distillation column are normally at their bubble point and vapours at their dew point, to reduce the temperature at which separation takes place we would reduce pressure. Thus operating at a lower pressure makes separation easier.

12.3 McCabe-Thiele Diagram

The technique usually taught to explore column operation is the *McCabe-Thiele Diagram*. While its use in industry is about as rare as the use of control theory, it does help us understand column operation. Figure 12.6 shows part of the McCabe-Thiele construction for the separation of propane (C_3) from butane (C_4). For a binary mixture, from Equation (12.9)

$$\alpha_{LK} = \frac{y}{x} \qquad \text{and} \qquad \alpha_{HK} = \frac{1-y}{1-x} \tag{12.11}$$

From Equation (12.10) relative volatility is given by

$$\alpha = \frac{\alpha_{LK}}{\alpha_{HK}} = \frac{y(1-x)}{x(1-y)} \tag{12.12}$$

Rearranging gives the equation for the vapour line

$$y = \frac{\alpha.x}{1+(\alpha-1)x} \tag{12.13}$$

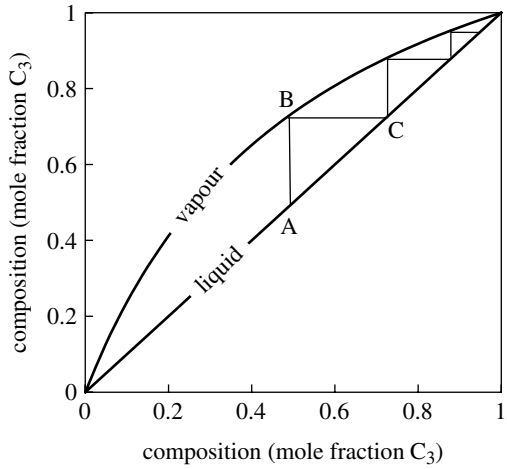

Figure 12.6 *Theoretical tray for propane/butane separation*

If we were to start with a liquid containing 0.5 mole fraction C_3 (point A) and allow it to partially evaporate in a closed container to equilibrium conditions, then the vapour would have the composition at B. If we were to remove the vapour and condense it, it would then contain about 0.75 mole fraction of C_3. The liquid left in the container would of course be substantially richer in C_4. This is one theoretical tray.

We can see that we only need repeat this process with the condensed vapour three or four more times to obtain a high purity C_3 product. Strictly we should recalculate α at each tray since, as shown by Figure 12.5, relative volatility changes with temperature and temperature reduces as we move up the column.

The method as shown assumes total reflux, i.e. no overhead product is drawn. The reflux ratio (R/D) is therefore infinite. It therefore gives us the minimum number of theoretical trays.

Figure 12.7 shows the molar balance for the basic distillation column on which most of this chapter is based. The molar feed rate to the column is F. It has *feed quality* of q. This rather confusing term has nothing to do with the composition of the feed but is a measure of its enthalpy. Its definition is given by

$$q = \frac{\text{heat used converting a mole of feed to saturated vapour}}{\text{latent heat of vaporisation of feed}} \tag{12.14}$$

But perhaps an easier way of understanding it is to think of it as the fraction of the feed which leaves the feed tray as liquid. Under normal conditions the liquid held on each tray is at its bubble point and the vapour leaving each tray is at its dew point. While this may not be true in the feed section, heat exchange between liquid and vapour takes place so that, within a few trays of the feed tray, equilibrium is reached.

However, if the feed entering is liquid exactly at its bubble point, no heat exchange is necessary. Thus all of the liquid entering as feed will leave the feed tray as liquid and so q will be 1. Similarly, if the feed entering is vapour exactly at its dew point, all of the feed will leave the feed tray as vapour and q is 0. Values of q between 0 and 1 are possible. Under these circumstances the feed will be a mixture of vapour and liquid at the saturation temperature. In fact q is then a measure of the *wetness* of the mixture.

Values outside the range of 0 to 1 are also possible. It is common for the feed to be liquid below its bubble point. On entering the column, the temperature of this liquid has to be raised to its bubble point. The energy to do this comes from condensing some of the vapour that would otherwise leave the feed tray. Thus the flow of

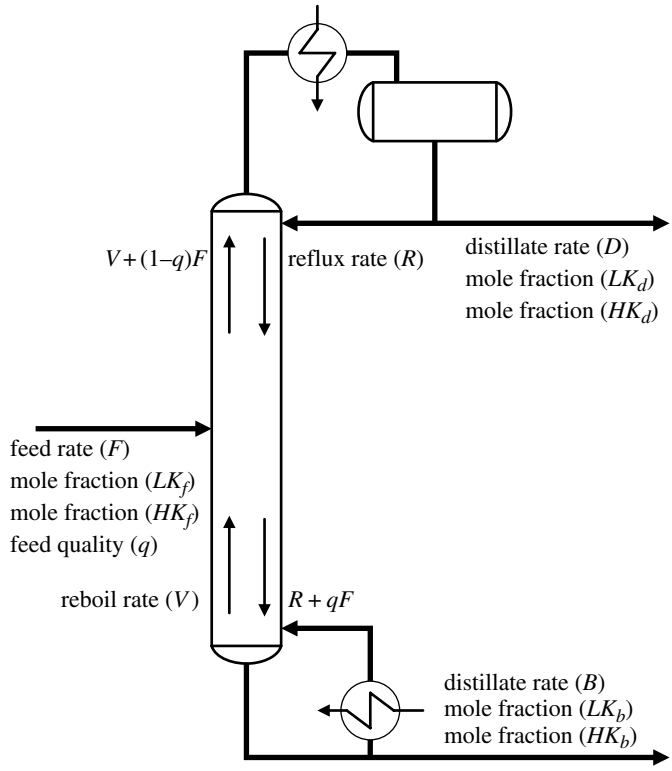

Figure 12.7 *Basic distillation column molar balance*

liquid leaving the feed tray will be greater than the flow of feed, so q is greater than 1. Similarly, if the feed is superheated vapour, on entering the column it gives up heat to bring its temperature down to dew point. This heat will vaporise some of the liquid that would otherwise have left the feed tray and so q is negative.

Trouton's Law states that, for a pure component

$$\frac{MW.\lambda}{T_b} = \text{constant} \tag{12.15}$$

MW is the molecular weight, λ the latent heat of vaporisation (per unit mass) and T_b is the boiling point (absolute temperature). To illustrate this law a number of components are included as Table 12.4. While not applicable over a wide range of components, the law is correct (to an accuracy of about 1%) for those that are closely related.

In distillation the boiling points of components will be similar – particularly when measured on an absolute basis. This means that the number of moles of liquid being evaporated is equal to the number of moles of vapour being condensed to provide the energy necessary – no matter what the composition of the vapour and liquid. This is known as *constant molal overflow* (or *equimolal overflow*) and means the molar liquid flow leaving each tray is the same for all the trays above the feed tray. While it changes to a new value at the feed tray it will be this value for all trays below the feed tray. The same can be said of the vapour flows.

Table 12.4 *Validation of Trouton's Law*

Component	MW	λ (kJ/kg)	T_b (K)	$\frac{MW.\lambda}{T_b}$
propane	44.10	425.68	231.0	81.3
propene	42.08	439.43	225.4	82.0
n-butane	58.12	385.95	272.7	82.3
water	18.02	2256.92	373.2	108.9
ethanol	46.07	840.54	349.5	110.8

We can therefore choose any tray in the upper section of the column and write the same mass balance. If x is the mole fraction of light key component in the liquid flowing on to a tray above the feed tray, and y is the mole fraction in the vapour leaving the same tray, then the balance around the top section of the column for the light key component, assuming a total condenser, is given by

$$\left[V+(1-q)F\right]y = R.x + D.LK_d \tag{12.16}$$

A mass balance round the top section of the column gives

$$V+(1-q)F = R+D \tag{12.17}$$

By combining Equations (12.16) and (12.17) we get the equation of the *top operating line*

$$y = \frac{R}{R+D}x + \frac{D.LK_d}{R+D} \tag{12.18}$$

We can similarly write a balance for the light key component for the section of the column below the feed tray.

$$(R+q.F)x = V.y + B.LK_b \tag{12.19}$$

From the overall mass balance

$$B = F - D \tag{12.20}$$

Rearranging Equation (12.17) gives

$$V = R + D - (1-q)F \tag{12.21}$$

Using Equations (12.20) and (12.21) to eliminate B and V from Equation (12.19), we get the equation of the *bottom operating line*

$$y = \frac{R+qF}{R+D-(1-q)F}x - \frac{(F-D)LK_b}{R+D-(1-q)F} \tag{12.22}$$

Provided the reflux is returned to the column at its bubble point, i.e. there is no sub-cooling by the condenser, Equations (12.18) and (12.22) can be plotted on the McCabe-Thiele diagram. These are shown

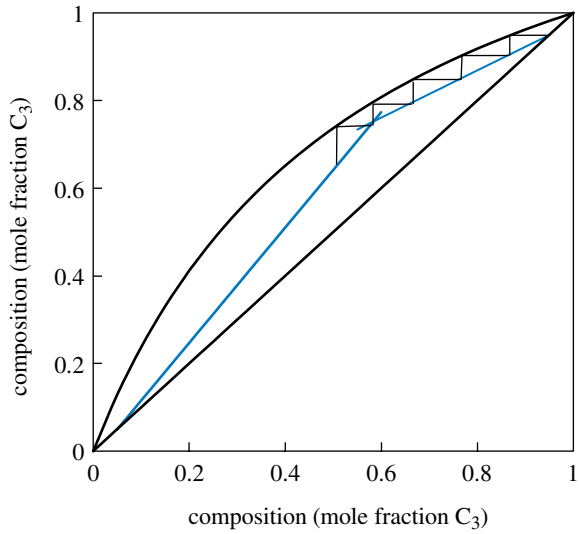

Figure 12.8 *Use of reflux for propane/butane separation*

in Figure 12.8; they lie between the vapour and liquid lines. By establishing a realistic (non-infinite) reflux we have increased the number of theoretical trays required. We have saved energy but now require a taller column. Column design is a trade-off between operating cost and cost of construction.

The number of actual trays will be greater than the theoretical number to allow for inefficiency. In reality the vapour and liquid leaving the tray will not be in equilibrium. To do so would require a very large tray residence time and an uneconomically large column diameter.

If we draw the McCabe-Thiele drawing for the separation of propene and propane (Figure 12.9) we can see that we will need far more theoretical trays. However, the propene/propane mixture is still 'well

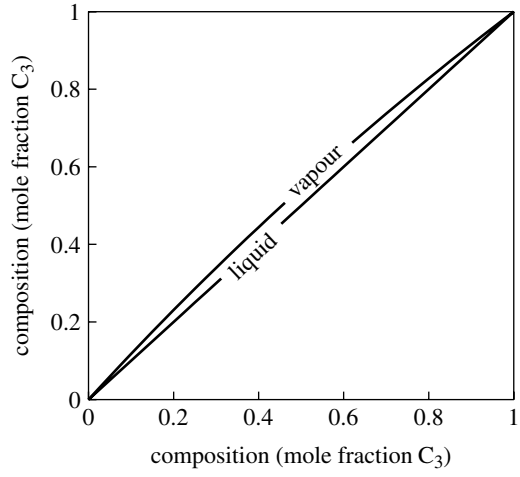

Figure 12.9 *Propene/propane separation*

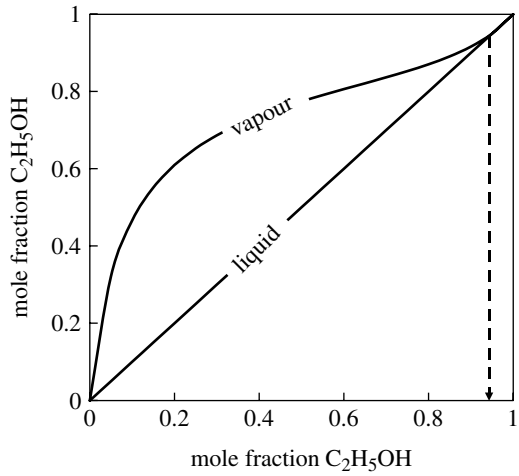

Figure 12.10 *Separation showing azeotrope*

behaved'. Figure 12.10 shows the vapour and liquid lines for the separation of ethanol and water. In this case the vapour line crosses the liquid line. Known as an *azeotrope*, it prevents the purity of ethanol exceeding around 95%. Indeed, if the feed contained a higher proportion of ethanol, the overhead product would be richer in water. The position of the azeotrope depends on pressure, so varying this may offer a solution. If not, then another component can be introduced. For example, if very high purity ethanol is required, the addition of benzene will permit this.

12.4 Cut and Separation

Cut and *separation* are the key parameters in determining the composition of the distillate and bottoms. We can write a total mass balance for the column in Figure 12.7.

$$F = D + B \tag{12.23}$$

We can also write a mass balance for one of the components, e.g. *LK*

$$F.LK_f = D.LK_d + B.LK_b \tag{12.24}$$

Combining these two equations to eliminate B gives

$$\frac{D}{F} = \frac{LK_f - LK_b}{LK_d - LK_b} \tag{12.25}$$

The proportion of feed drawn as distillate is the *distillate cut*. We could equally have developed an expression for *bottoms cut*.

$$\frac{B}{F} = \frac{LK_d - LK_f}{LK_d - LK_b} = 1 - \frac{D}{F} \tag{12.26}$$

If we take the example in Table 12.1

$$\frac{D}{F} = \frac{45-10}{95-10} = 41\% \tag{12.27}$$

This shows us that 41% of the feed must be drawn as distillate. This is a necessary condition to exactly meet both product composition targets. Depending on how the composition targets are defined we may need to refine the calculation for the required cut. For example, defining the specification on distillate in terms of C_4 content (HK_d) is equivalent to setting a target for LK_d only if it is truly a binary distillation. For non-key components we can apply the *Hengstebeck approximation* – which assumes that there is no LLK in the bottoms and no HHK in the distillate. While usually not exactly the case this enables us to write Equations (12.28) and (12.29) to illustrate the impact of a multi-component feed.

$$F.LLK_f = D.LLK_d \tag{12.28}$$

$$HK_d + LK_d + LLK_d = 100 \tag{12.29}$$

Combining these with Equations (12.23) and (12.24) gives

$$\frac{D}{F} = \frac{LK_f - LK_b + LLK_f}{100 - HK_d - LK_b} \tag{12.30}$$

So if the distillate specification is defined as a maximum impurity (rather than a minimum purity) and there were any change in LLK_f, it would be necessary to change the cut. As Equation (12.31) shows, the same would apply if the bottoms specification is expressed as a maximum impurity and there is a change in HHK_f.

$$\frac{D}{F} = \frac{100 - HK_f - LK_b - HHK_f}{100 - HK_d - LK_b} \tag{12.31}$$

We conclude that any control scheme we design should only change the distillate flow if the feed rate changes (to maintain *D/F* constant), if the feed composition changes or if there is any change in the target composition for either product. If distillate flow is changed for other reasons, such as a change in column pressure or feed enthalpy, then it is certain that at least one product composition will move away from target.

Rearranging Equation (12.25)

$$LK_b = \frac{LK_f - \left(\frac{D}{F}\right)LK_d}{1 - \left(\frac{D}{F}\right)} \tag{12.32}$$

Putting in the values from our example gives

$$LK_b = \frac{45 - 0.41LK_d}{1 - 0.41} = 76.5 - 0.697LK_d \tag{12.33}$$

The values of 10% for LK_b and 95% for LK_d are only one solution of this equation. In fact there is a wide range of solutions as shown in Figure 12.11 by the coloured line. This shows that, if the cut is wrong, we cannot meet both product specifications. For example if D/F is fixed at 30% we can meet the target for LK_d but not that for LK_b. If it is set at 50% we can meet LK_b but not that for LK_d. Meeting the required cut is a necessary but not a sufficient condition of meeting the target compositions. For example $LK_d = LK_b = LK_f$ is one solution to Equation (12.33); the cut may be correct but there is no separation. Similarly a value of 75% for LK_d and 24% for LK_b satisfies Equation (12.33). This reduced separation example is shown in Figure 12.11.

We can also show diagrammatically (Figure 12.12) the effect of changing separation. In both the target and reduced separation cases the distillate cut is maintained at 41%. For multi-component distillation we can use a *true boiling point (TBP) curve* to represent the same two cases. A TBP curve is obtained in the laboratory, in principle, by slowly heating the liquid sample and recording the volume evaporated. In practice the test is a little more complex. Originally it was performed as a batch distillation using a column

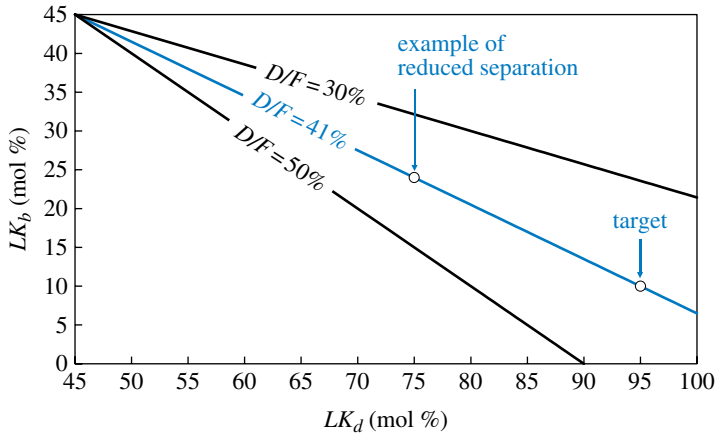

Figure 12.11 *Importance of cut*

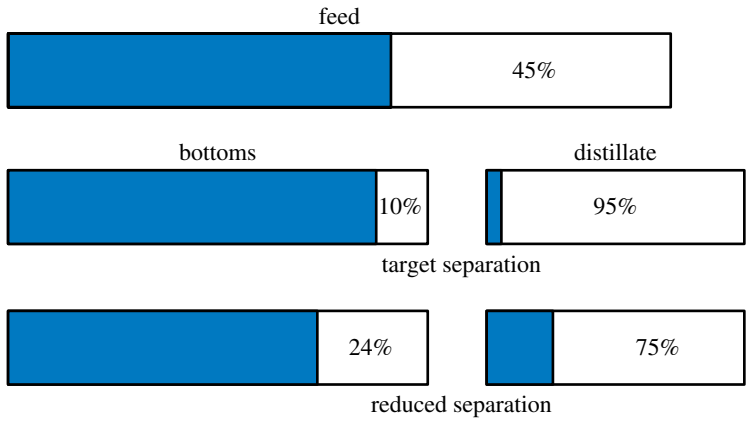

Figure 12.12 *Changing separation*

with a very large number of trays and a very high reflux ratio. Nowadays it is done by mass spectrometry. Figure 12.13 shows the curve for the feed to our case study column. IBP is the *initial boiling point* and FBP the *final boiling point* (or *end point*).

Our cut-point, currently expressed as a fraction of feed, can now be defined as a *cut-point temperature*. In concept, any feed material boiling at a temperature below this value will leave the column as distillate and any boiling above this temperature will leave as bottoms. In practice such a perfect separation is impossible. Figure 12.14 shows the TBP curves for the two products if we meet our target separation. It shows that, as required, 95% of the distillate boils below the cut-point – as does 10% of the bottoms. As a result, the TBP curves overlap. This overlap is a measure of separation and is defined as the FBP of the distillate less the IBP of the bottoms. Figure 12.15 shows the reduced separation case where the overlap is far greater. Table 12.5 shows how the compositions in Table 12.1 are modified by changing to the reduced separation.

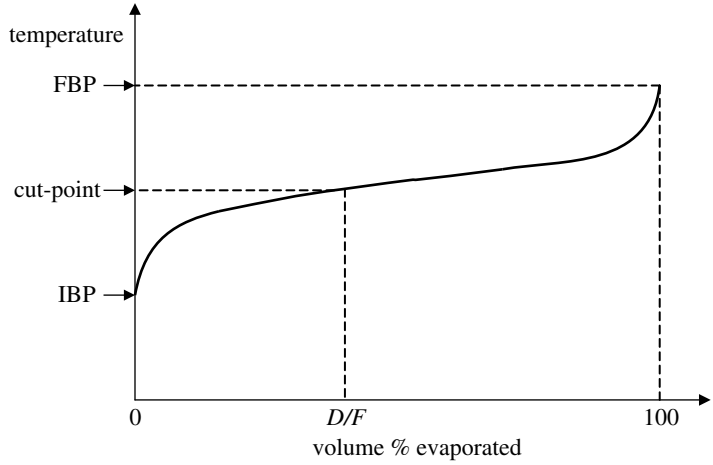

Figure 12.13 *TBP curve for column feed*

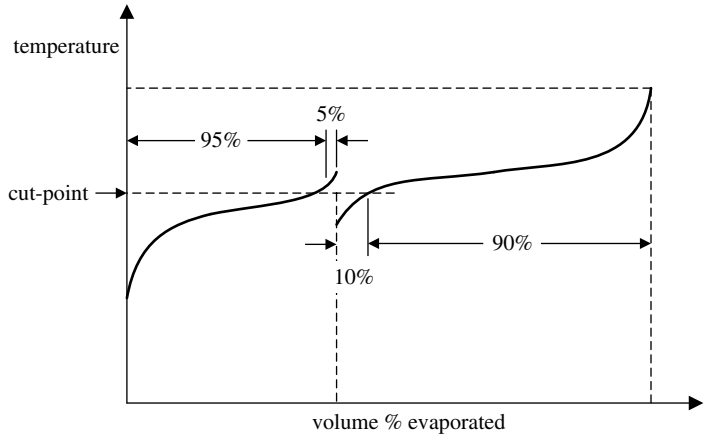

Figure 12.14 *Target separation*

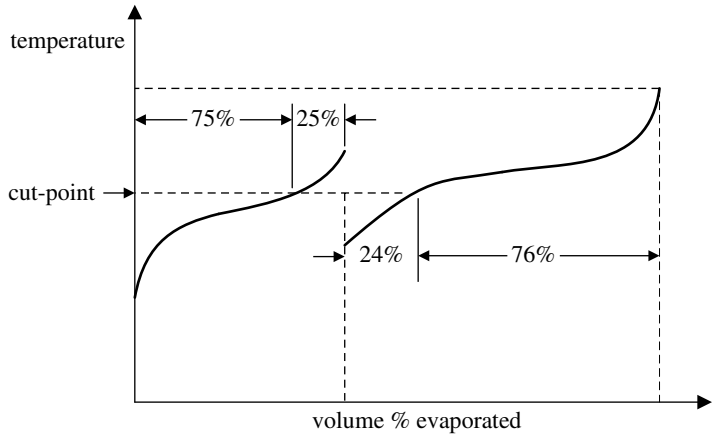

Figure 12.15 *Reduced separation*

Table 12.5 *Reduced separation case*

Component	Feed	Distillate		Bottoms	
	moles/hr	moles/hr	mol %	moles/hr	mol %
C_2H_6	0.7	0.5	1.3	0.2	0.4
C_3H_8	45.0	30.9	75.0	14.1	24.0
C_4H_{10}	52.7	9.6	23.5	43.1	73.1
C_5H_{12}	1.6	0.1	0.2	1.5	2.5
Total	100.0	41.1	100.0	58.9	100.0

Figure 12.16 is another way, used by some, of illustrating the effect of cut and separation. The solid coloured line shows the target operation. The reboiler duty has been set to give the required separation. The 41% cut line intersects the HK_d line at 5% and the LK_b line at 10% – the target compositions. For the reduced separation case the reboiler duty has been reduced; the 41% cut line now intersects at 25% and 24%. Also shown, as the dashed line, is the 'perfect' separation case. Not attainable on a real column, it shows what the compositions would be if reboiler duty was unlimited.

In order to quantify separation (S) we will take the definition from the *Fenske Equation*.

$$\alpha^N = S = \frac{\left(\frac{LK_d}{HK_d}\right)}{\left(\frac{LK_b}{HK_b}\right)} \tag{12.34}$$

For multi-component distillation the value used for relative volatility (α) is the geometric mean of that at the top of the column (α_t) and that at the bottom (α_b), i.e. $(\alpha_t.\alpha_b)^{0.5}$. The Fenske Equation is normally used to estimate the minimum number of theoretical trays (N) for a column, i.e. the number necessary to achieve the required separation when operating with total reflux. While of little value in column design it is a parameter used by others in inferential property calculations. We will cover this later in this chapter.

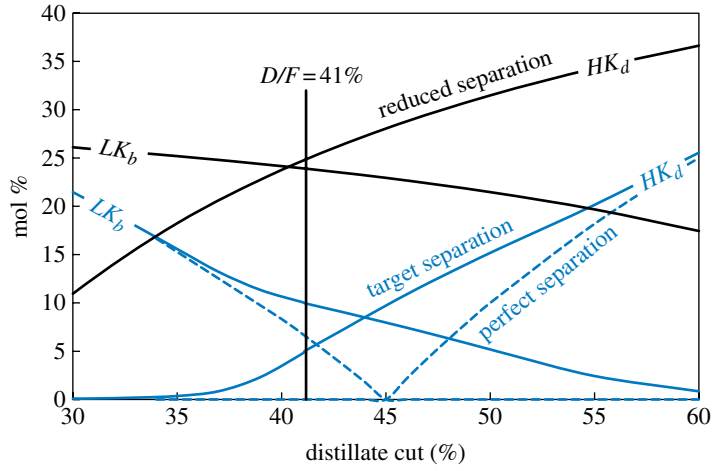

Figure 12.16 *Effect of cut and separation*

Referring to Table 12.1, we can calculate our target separation.

$$S = \frac{\left(\frac{95.0}{3.5}\right)}{\left(\frac{10.0}{87.1}\right)} = 236 \tag{12.35}$$

From Table 12.5 we can similarly calculate it for the reduced separation case.

$$S = \frac{\left(\frac{75.0}{23.5}\right)}{\left(\frac{24.0}{73.1}\right)} = 9.7 \tag{12.36}$$

We see that for a change of about 20% in product purity the value for S changes by two orders of magnitude. To obtain a more linear relationship, $\log(S)$ is commonly used. Indeed, the Fenske Equation is often written in the form:

$$N = \frac{\log(S)}{\log(\alpha)} \tag{12.37}$$

If there is no separation the distillate composition will the same as the bottoms composition and so S will be 1 and $\log(S)$ will be 0.

Equation (12.34) can be rewritten as

$$S = \frac{\left(\frac{LK_d}{100 - LK_d - LLK_d - HHK_d}\right)}{\left(\frac{LK_b}{100 - LK_b - LLK_b - HHK_b}\right)} \tag{12.38}$$

Rearranging

$$LK_b = \frac{\left(100 - LLK_b - HHK_b\right) LK_d}{\left(100 - LK_d - LLK_d - HHK_d\right) S + LK_d} \tag{12.39}$$

We can assume that, provided cut is kept constant, the concentrations of the off-key components in the products vary little as separation is changed. This allows us to add, to the line of constant cut (that we saw in Figure 12.11), lines of constant separation – as shown in Figure 12.17.

With some manipulation of the equations above it is possible to plot the same chart based on purity targets (i.e. LK_d versus HK_b) or impurity specifications (i.e. HK_d versus LK_b). For example, using the feed composition from Table 12.1, setting the target for HK_d at 5 mol% and that for LK_b at 10 mol% and applying Equation (12.31) shows that the distillate cut must be 42%. Separation must be 163. The modified chart is included as Figure 12.18.

For truly binary systems the required S can be calculated from the product purity targets.

$$S = \frac{\left(\dfrac{LK_d}{100 - LK_d}\right)}{\left(\dfrac{100 - HK_b}{HK_b}\right)} = \frac{LK_d HK_b}{\left(100 - LK_d\right)\left(100 - HK_b\right)} \tag{12.40}$$

An alternative approach is to calculate the required S from the product impurity targets.

$$S = \frac{\left(\dfrac{100 - LK_b}{LK_b}\right)}{\left(\dfrac{HK_d}{100 - HK_d}\right)} = \frac{\left(100 - LK_b\right)\left(100 - HK_d\right)}{LK_b HK_d} \tag{12.41}$$

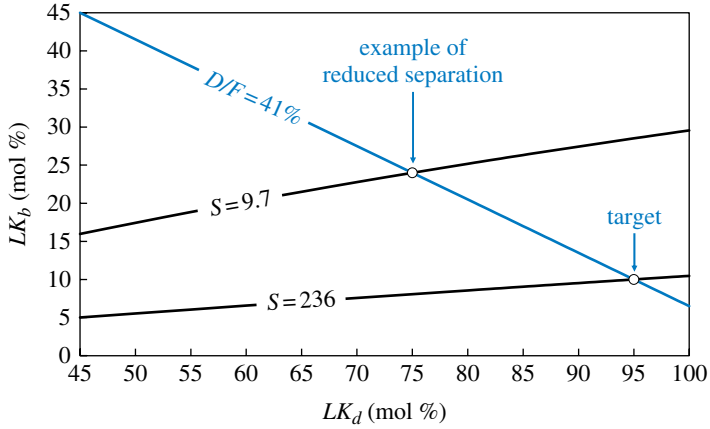

Figure 12.17 *Importance of separation*

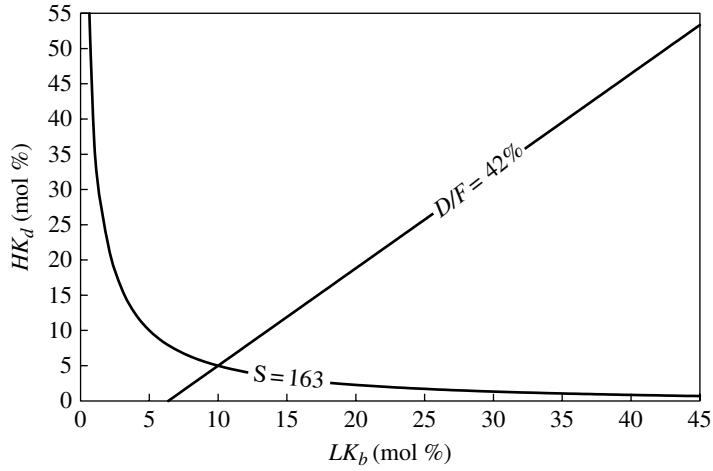

Figure 12.18 *Cut and separation required to meet impurity targets*

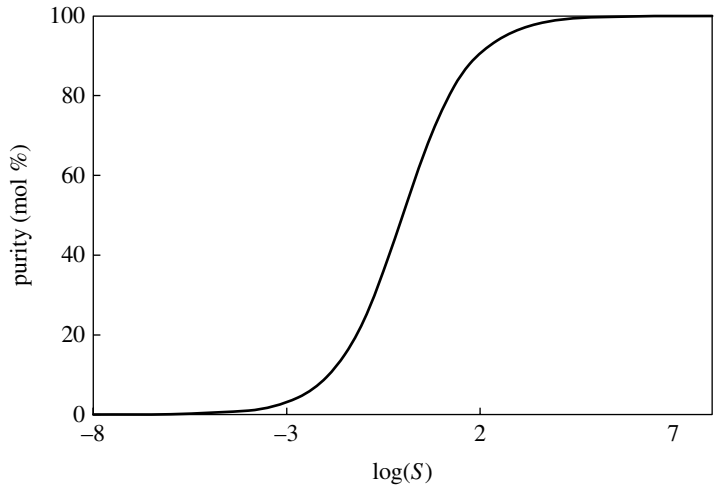

Figure 12.19 *Relationship between purity and separation*

Figure 12.19 shows that, despite using log(S), the relationship with purity remains nonlinear.

So far we have assumed that we want to exactly meet the specifications. However, targets are usually set as inequalities; for example our propane has to be at least 95% pure. It is permitted to produce it at a higher purity and, should there be an economic advantage, we might wish to do.

Let us consider first the situation where only the distillate product has a specification. We are permitted to produce a bottoms product of any composition. For example, if we chose to produce bottoms containing 20% propane, then our distillate cut, from Equation (12.25) would be

$$\frac{D}{F} = \frac{45-20}{95-20} = 33\% \tag{12.42}$$

While we are free to choose our own target for LK_b, it is still subject to constraints. For example, it cannot exceed LK_f. The bottoms cannot contain more light key material than the feed. Equation (12.25) would require us to operate with a negative cut. We can, however, reduce LK_b to zero. While it is not possible to make both products 100% pure, it is possible to make either product completely pure – even on columns with relatively poor separation. It may of course be economically disastrous in that doing so might lose very large quantities of the more valuable component in the lower value product.

Thus

$$0 < LK_b < LK_f \tag{12.43}$$

If we put these constraints into Equation (12.25) we get

$$\frac{45-45}{95-45} < \frac{D}{F} < \frac{45-0}{95-0} \quad \text{or} \quad 0\% < \frac{D}{F} < 47\% \tag{12.44}$$

The effect that cut has on the bottoms composition, while keeping the distillate composition constant, is shown in Figure 12.20. While it is possible, provided we keep the distillate cut between 0 and 47%, to meet the distillate 95% purity target, we have to compensate by adjusting separation. As shown in Figure 12.21, as cut approaches the upper limit, separation approaches infinity.

A similar approach allows us to develop the condition for keeping the bottoms composition at its specification. We now assume that LK_b is fixed at 10% but we are free to choose any target for LK_d. The distillate cannot contain less light key than the feed and so LK_d is constrained as

$$LK_f < LK_d < 100 \tag{12.45}$$

If we put these constraints into Equation (12.25) we get

$$\frac{45-10}{100-10} < \frac{D}{F} < \frac{45-10}{45-10} \quad \text{or} \quad 39\% < \frac{D}{F} < 100\% \tag{12.46}$$

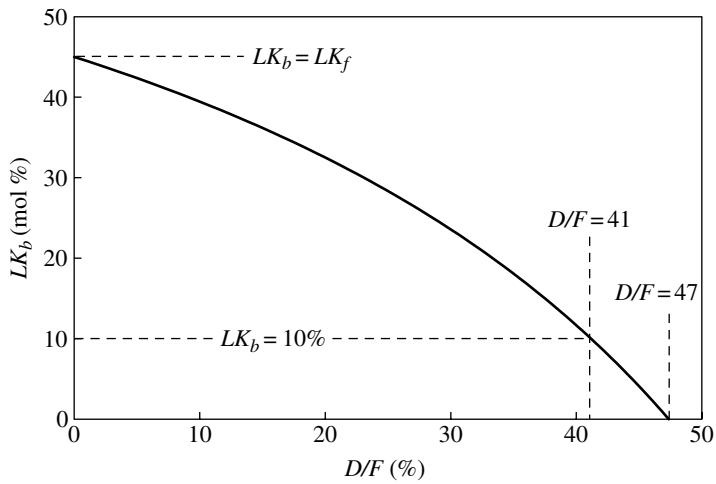

Figure 12.20 *Effect of cut on bottoms composition*

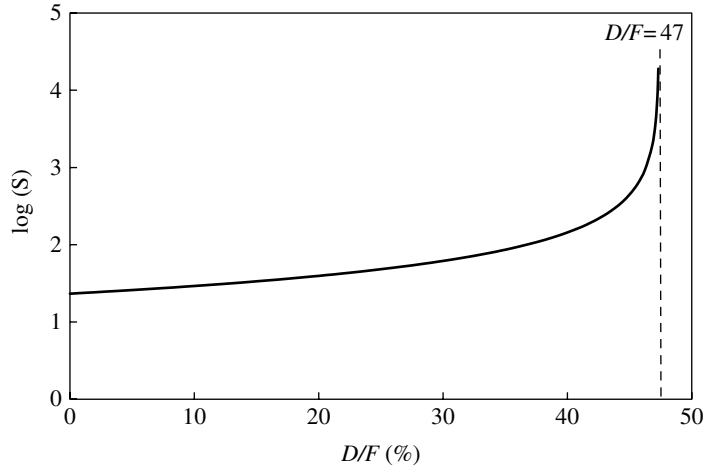

Figure 12.21 *Keeping distillate composition at target*

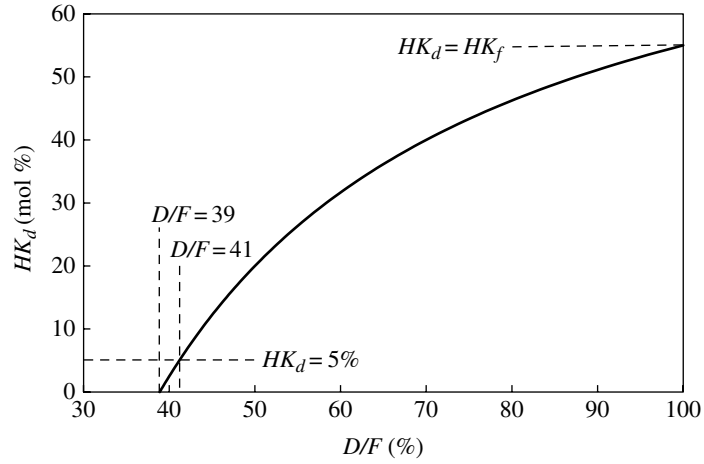

Figure 12.22 *Effect of cut on distillate composition*

The effect that cut has on the distillate composition, while keeping the bottoms composition constant, is shown in Figure 12.22. For simplicity we have treated the distillation as truly binary so that

$$HK_f = 100 - LK_f \tag{12.47}$$

$$HK_d = 100 - LK_d \tag{12.48}$$

$$HK_b = 100 - LK_b \tag{12.49}$$

While it is possible, provided we keep the distillate cut between 39 and 100%, to meet the bottoms 90% purity target, we have to compensate by adjusting separation. As shown in Figure 12.23, as cut approaches the lower limit, separation again approaches infinity.

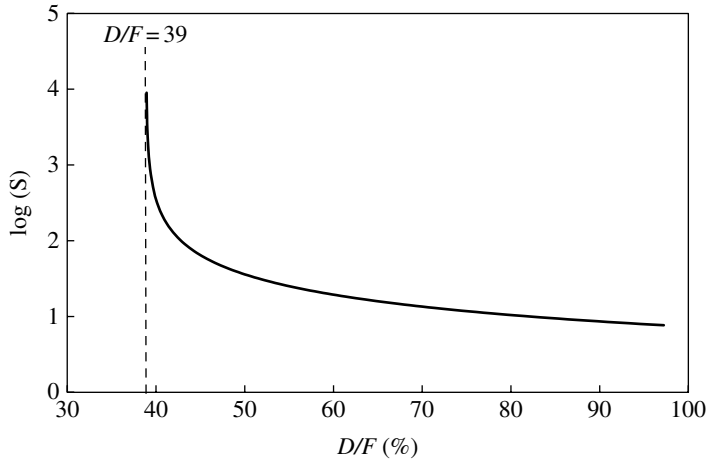

Figure 12.23 *Keeping bottoms composition at target*

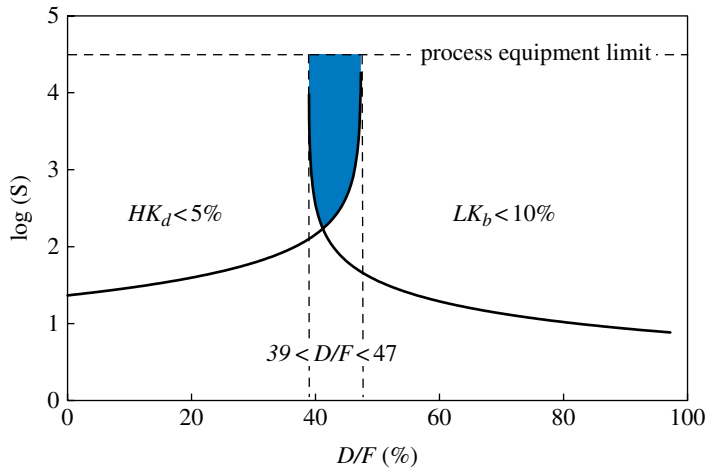

Figure 12.24 *Keeping both product compositions at target*

We now combine the curves from Figures 12.21 and 12.23 to show the constraints in keeping both products within specification. We must operate within the area coloured in Figure 12.24. This confirms that, if we require both product compositions to be exactly on target we must operate where the lines cross, with a distillate cut of 41%. This requires the lowest separation and therefore the lowest energy.

If we treat the specifications as constraints rather than absolute targets, then we can operate at a distillate cut within the range 39 to 47%. However, as we approach these limits, tray hydraulics or condenser/reboiler capacity will constrain how much separation can be increased. The true feasible range will be considerably narrower.

Where we wish to operate within this feasible space will depend on the process economics. If both products have a similar value, then there is no incentive to maximise one at the expense of the other. Profit would therefore be maximum by operating exactly on both specifications and would be reflected as an

energy saving. But, for example, if propane were considerably more valuable than butane, then we would not wish to leave C_3 material in the bottom product. It would be economic to produce butane at purity greater than the specification demands. We would, however, want to operate at the lowest permitted propane purity in order to maximise the amount of C_4 material that is sold at the propane price. We would therefore wish to operate at a higher distillate cut; how large depends on the cost of the additional energy required to recover the additional propane. This falls into the area of optimisation that we cover later in this chapter.

While the feasible range of around 8% for distillate cut might seem quite broad, this is as a result of the purity targets being relatively low. If, for example, both products were required to be better than 99.5% pure, then the feasible range would be given by Equation (12.25) as

$$\frac{45-0.5}{100-0.5} < \frac{D}{F} < \frac{45-0}{99.5-0} \qquad \text{or} \qquad 44.7\% < \frac{D}{F} < 45.2\% \tag{12.50}$$

Such purity targets are common for propene destined for polymer production, adding a further challenge to the already difficult separation from propane.

In practice we do not have separation as a MV; we instead adjust it by changing *fractionation*. The terms separation and fractionation are often used interchangeably. This is not strictly correct. Separation is a measure of product composition, while fractionation is a measure of energy used. They are certainly related in that increasing fractionation will increase separation but, as we will see later, separation is influenced by many other parameters.

Figure 12.25 shows the effect that varying the reboiler duty has on the composition of both products. In this case cut has been kept constant which, depending on how the basic controls have been configured, may not be the real situation. It nevertheless shows that, as we change fractionation, both product compositions change.

Similarly, depending on the control configuration, we may not be able to directly manipulate distillate flow rate and keep fractionation constant. However, for the purposes of explanation, this has been the test performed for Figure 12.26. We see that changing it again affects both compositions.

These figures illustrate some of the issues that our control design will later have to address. Firstly, if both MVs affect both product compositions, which one do we select for distillate composition control and

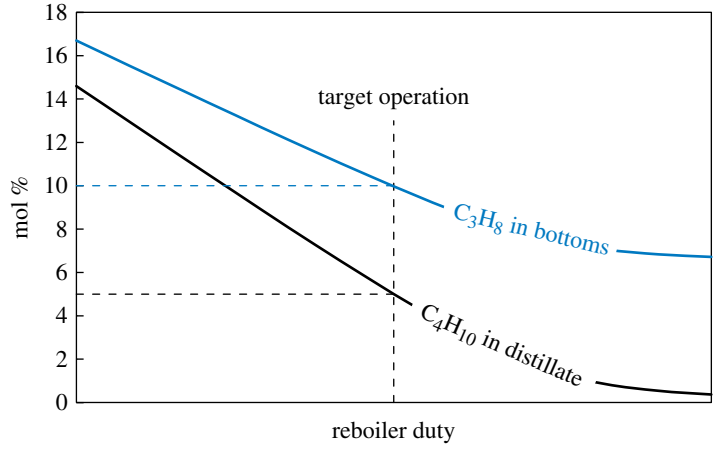

Figure 12.25 *Effect of reboiler duty on product compositions*

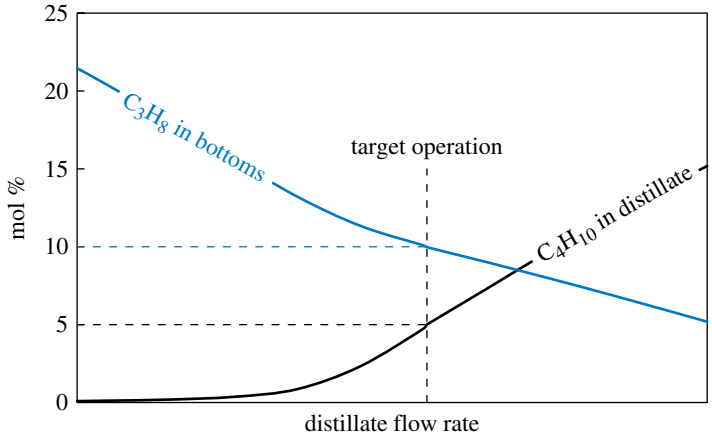

Figure 12.26 *Effect of distillate flow rate on product compositions*

which for bottoms? Secondly, if we adjust one MV to correct off-grade production, how do we deal with the problem that this will put the other product off grade? Further, while the relationships between composition and the MVs are almost linear in the region in which we wish to operate, how would we handle the much more nonlinear behaviour if we were to operate with much higher purities?

12.5 Effect of Process Design

Separation is determined by a number of factors – both operating conditions and column design. Figure 12.27 shows the effect of varying reboiler duty while keeping the cut constant. Under these circumstances increasing duty has no effect on product yields. But it does increase the liquid and vapour rates in the column and so improves the purity of both products and hence separation. The relationship is highly nonlinear; as reboiler duty is increased the benefit of the additional duty reduces. The achievable purity is limited by the combination of the number of trays and their hydraulic limitations – as previously described in Figure 12.3.

Figure 12.28 shows the effect of varying the number of trays, again with the cut held constant. Again we reach the point where the addition of further trays gives no benefit. Separation is limited by the reboiler duty. Figure 12.29 combines the effect of reboiler duty with that of the number of trays. By fixing the product compositions we can assess the energy saving as the number of trays is increased. As expected from the previous figures the relationship is highly nonlinear. Too few trays mean that the column would be extremely costly to operate. Too many greatly increase construction cost with virtually no impact on energy savings.

In much the same way that separation is affected by the number of trays in the column, it is also affected by their efficiency. Inefficiencies arise because of incomplete mixing on each tray and because it is not practical for the residence time to be large enough for the vapour and liquid to reach equilibrium. There are two common definitions of tray efficiency (η). The *Fenske tray efficiency* is applied to the whole column.

$$\eta = \frac{N_{theoretical}}{N_{actual}} \tag{12.51}$$

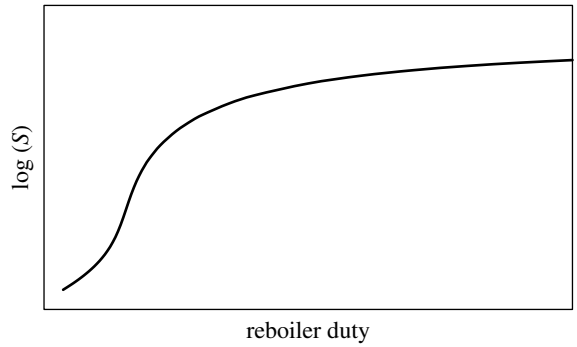

Figure 12.27 *Effect of reboiler duty on separation*

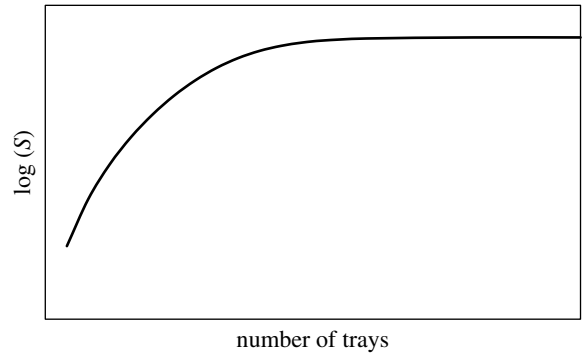

Figure 12.28 *Effect of number of trays on separation*

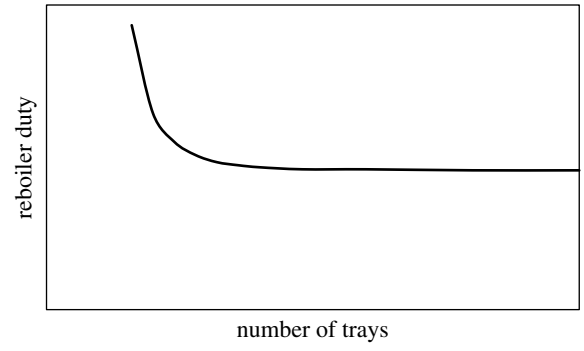

Figure 12.29 *Trade-off between number of trays and reboiler duty*

We can use Equation 12.51 to redefine the relative volatility for use in the McCabe-Thiele diagram.

$$\alpha_{actual} = \left(\alpha_{theoretical}\right)^{\eta} \tag{12.52}$$

The *Murphree tray efficiency* is defined as the fraction of the theoretical change in composition actually achieved across a tray; for the n^{th} tray.

$$\eta_n = \frac{\left(y_n\right)_{actual} - y_{n-1}}{\left(y_n\right)_{theoretical} - y_{n-1}} \tag{12.53}$$

To incorporate this definition into the McCabe-Thiele diagram, we plot an actual vapour line part way between the operating lines and the theoretical vapour line.

Figure 12.30 shows the effect that tray efficiency has on separation. As we did when we varied the number of trays we can explore what change in reboiler duty is necessary to maintain the compositions constant as tray efficiency changes. This is shown in Figure 12.31. Tray efficiency varies widely depending on the material being processed. *O'Connell's correlation* [28] predicts tray efficiency based on relative volatility (α) and viscosity at the operating temperature (μ cP). It was developed empirically for bubble cap columns. For sieve trays and valve trays it underestimates efficiency.

$$\eta = 0.503\left(\alpha.\mu\right)^{-0.226} \tag{12.54}$$

To predict the Fenske efficiency, we would use the molal average feed viscosity at the average column temperature. For LPG this is typically 0.13 cP. Assuming a typical value of 3 for relative volatility (from Figure 12.5) gives a predicted efficiency of 62% for a LPG splitter. For a propene/propane splitter, with a relative volatility of about 1.2, it would be about 77%. As expected these are lower than the more typical value of around 85% achievable with other trays but, as Figure 12.31 shows, there is little to be gained in energy savings by upgrading to a more efficient tray. However, columns processing much heavier components derived from crude oil, such as bitumen, would have an efficiency of around 20% and offer the potential for large energy savings.

The location of the feed tray also affects separation. Figure 12.32 shows the effect of moving this above and below its optimum location. Again we can adjust the reboiler duty to compensate for poor positioning of the feed tray. Figure 12.33 shows the result when maintaining constant product compositions

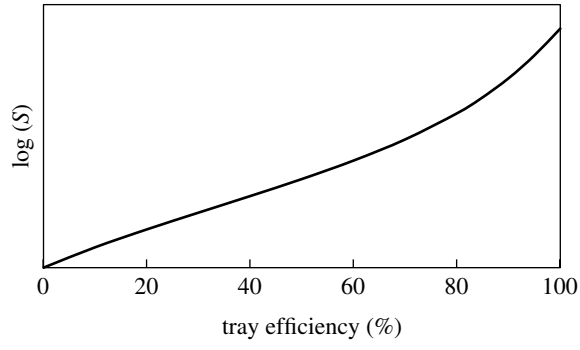

Figure 12.30 *Effect of tray efficiency on separation*

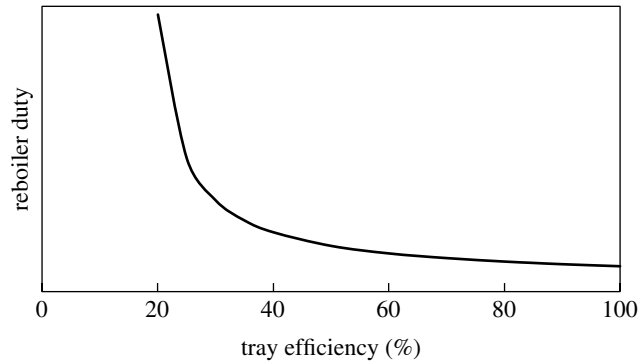

Figure 12.31 *Effect of tray efficiency on reboiler duty*

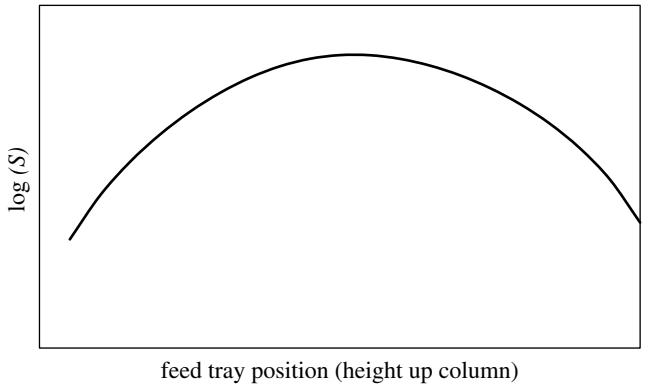

Figure 12.32 *Effect of feed tray location on separation*

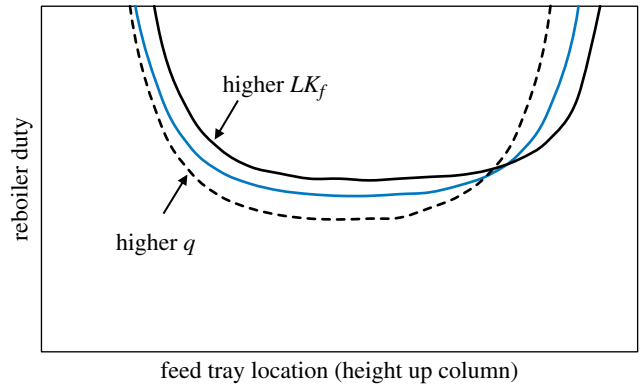

Figure 12.33 *Optimising feed tray location*

(coloured line). Sensitivity studies also show the effect of changing the feed composition and feed enthalpy. In this case the optimum is fairly 'flat' and so the cost of being a tray or two away from the ideal location is small. It is common to build flexibility into the column design and include the facility to switch feed tray location. Switching is done by operation of manual block valves and is not something that would normally be included in a control strategy – indeed, switching would normally be performed with the unit shut down. As part of control design, the control engineer might therefore wish to confirm that the optimum feed tray location (N_{feed}) has been chosen – using, for example, the *Kirkbride Equation*. The equation can be applied to either theoretical or real trays. If trays are numbered from the bottom of the column then

$$N_{feed} = \frac{N}{1+\left[\frac{B.HK_f}{D.LK_f}\left(\frac{LK_b}{HK_d}\right)^2\right]^{0.206}} \tag{12.55}$$

Combining Equations (12.23) and (12.25) to eliminate F gives

$$\frac{B}{D} = \frac{LK_d - LK_f}{LK_f - LK_b} \tag{12.56}$$

Substituting in Equation (12.55) gives

$$N_{feed} = \frac{N}{1+\left[\frac{\left(LK_d - LK_f\right)HK_f LK_b^2}{\left(LK_f - LK_b\right)LK_f HK_d^2}\right]^{0.206}} \tag{12.57}$$

Using the product compositions in Table 12.1 and assuming the column has 20 trays, Figure 12.34 shows the application of this equation (using our target product compositions) to show the effect of feed composition. It confirms, in this case, what Figure 12.33 shows. Quite large changes would be necessary to justify

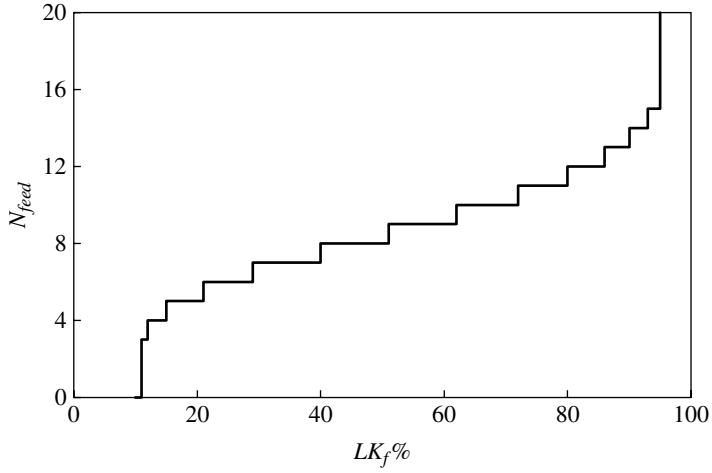

Figure 12.34 *Effect of feed composition on optimum feed tray location*

relocating the feed tray. Similar calculations can be made to determine whether a change in product composition targets would have a significant impact.

12.6 Basic Controls

There are several issues which need to be addressed when designing the basic controls for the column. The first is one of *pairing*. We will see that on our simple column that there are five PVs that we must control – pressure, reflux drum level, column base level, distillate composition and bottoms composition. We normally have available five MVs – distillate flow, bottoms flow, reflux flow, reboiler duty and condenser duty. We need therefore to decide which MV is going to be used to control which PV. Theoretically there are 120 ($5 \times 4 \times 3 \times 2$) possible combinations. While many of these are nonsensical, a large number of feasible schemes are possible.

Our next problem is one of nonlinearity. We have seen already that the relationship between separation and fractionation can be highly nonlinear. We can also expect the process dynamics to be difficult. Parts of the process, such as the reflux drum, hold large inventories that will introduce large process lags. Each tray introduces a transport delay – resulting in a very large process deadtime on columns with a large number of trays. Dynamics can vary; for example, at higher internal reflux flows, the liquid level on the trays increases – so increasing the overall column inventory and hence the lag. Further we have already seen the interactive nature of the process – changing any one of the MVs affects all of the PVs.

The process will be subject to multiple disturbances. Often fed from an upstream unit the feed rate, feed composition and feed enthalpy can all vary. Disturbances may enter through the reboiler – particularly if the duty is provided through heat integration with another part of the process. Disturbances can enter through the condenser – particularly from *air-fin* types that are subject to sudden changes in ambient temperature or rainfall. We may deliberately vary pressure to optimise the column and we may occasionally change product specifications.

Some of the instrumentation, most notably on-stream analysers, can be very costly. Chromatographs are commonly used which, along with the necessary housing and sampling system, are particularly costly. Other types such as near-infra red (NIR) and nuclear mass resonance (NMR) devices can be more so and involve high ongoing support costs. It may be that we cannot economically justify such instrumentation and need to compromise on the control design.

The basic instrumentation available to us is shown in Figure 12.35. For our case study we have assumed that the condenser uses cooling water and the reboiler uses a heating fluid, for example, steam. We will examine alternatives later. While the drawing does show a gaseous product, this is downgraded to flare and is only intended to be generated when required to avoid over-pressurisation.

By most standards the column is generously instrumented – particularly with on-stream analysers. It is not the intention to suggest that all columns should be so endowed. Not all of the schemes we will design for this column are economically justifiable or applicable on all. It is assumed by this stage the control engineer needs no help in installing the necessary flow controllers so these have already been included on the products, reflux and reboiler heating fluid.

12.7 Pressure Control

One of the key requirements of the basic column controllers is to maintain the energy balance. Energy enters the column as feed enthalpy and in the reboiler. It leaves as product enthalpy and in the condenser. If we neglect losses these inputs and outputs must balance. While we may have some limited control over

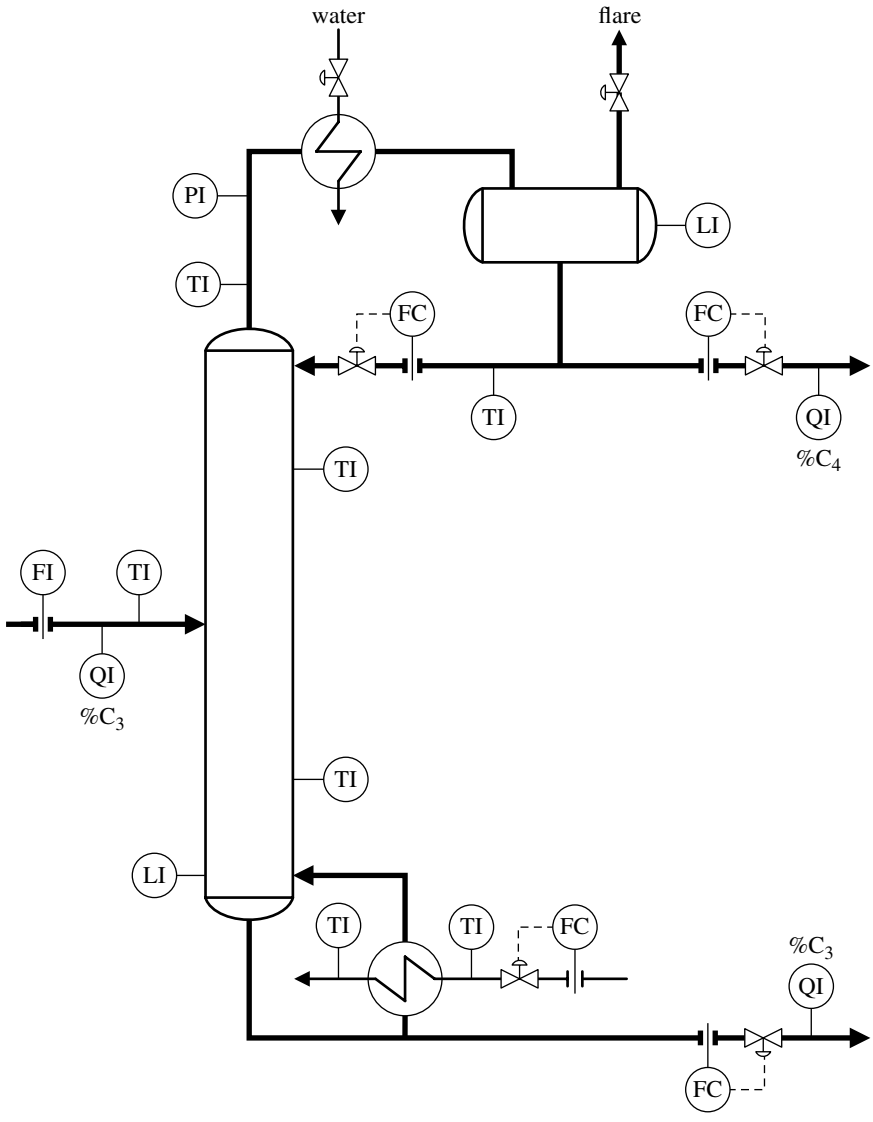

Figure 12.35 *Basic instrumentation*

feed enthalpy, and maybe some control over product enthalpy, the main source of energy is the reboiler and the main sink is the condenser. If the input energy is greater than the output, then more vapour will be produced than condensed and the column pressure will rise. By controlling column pressure we therefore maintain the energy balance.

Pressure makes for good control because it responds quickly to any energy imbalance, but there are also other good reasons for controlling it. Clearly we wish to ensure the unit is safe and that normal process disturbances do not lift relief valves. Pressure affects tray loading. If pressure is reduced, given that the space occupied by the vapour is constant, the vapour velocity must increase to maintain the same molar

flow. If operating close to its design limit the disturbance could result in the column hitting the blowing or flooding constraint. Similarly pressure affects dew points and bubble points which in turn change the log mean temperature difference (LMTD) across the condenser and across the reboiler. Again, if operating close to design limits, one of these may become a temporary capacity limit.

We have seen that pressure affects bubble points and relative volatility; any disturbance to pressure will therefore disturb product composition. While we will show later that there can be advantages in adjusting pressure, we want to do this in managed way, making the changes slowly and taking compensating action to maintain constant composition. We want to avoid deviations from the pressure controller SP.

There is a very wide range of possible process designs for the control of pressure. It is beyond the scope of this book to go into great detail of the process aspects. Its aim is to highlight the issues that can arise with controller design and tuning. The schemes fall into three fundamental groups – those that adjust the rate at which vapour is condensed, those that adjust the rate at which vapour leaves the process and those that adjust the vapour generated.

Taking the first of these groups, there are a number of possible methods – including manipulation of coolant rate or temperature, changing the efficiency of the condenser or partially bypassing it.

Figure 12.36 shows one of the most common schemes – manipulation of coolant flow. Commonly the coolant will be water which may come from a closed system where heat is removed from the circulating water in a cooling tower. Or it may be a once-through system taking water from the sea or a river. A flow controller on the water is normally not justified. If the water supply pressure is reasonably constant it offers no advantage and, for natural sources of water, may be prone to fouling and corrosion. There is often a minimum limit put on the water flow. In fouling services it is important to keep the flow turbulent to keep any solids in suspension. In salt water service it is often necessary to keep the water exit temperature below about 50°C (around 120°F), to prevent excessive corrosion of the mild steel tube bundle. The response of the controller can be highly nonlinear. At high coolant flow rates the coolant exit temperature approaches the supply temperature and further increases in flow have little impact on heat transfer. Figure 12.37 illustrates this. If the pressure controller directly manipulates the valve the nonlinearity can give tuning issues. As described in Chapter 5, an equal percentage valve can help alleviate the problem. In any case, because of the relatively fast process dynamics, the potential sluggish response when the process gain is low generally goes unnoticed.

On columns where the temperature of the coolant must be lower than that possible using water a refrigerant must be used. This might be ammonia or propane. Figure 12.38 shows a typical configuration. The

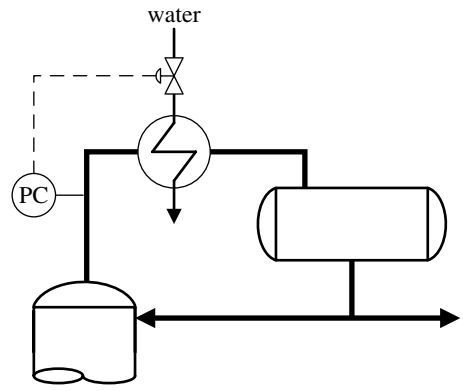

Figure 12.36 *Manipulating cooling water*

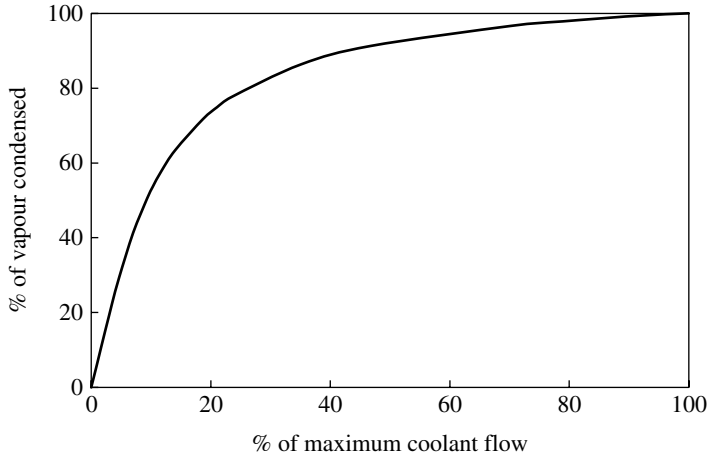

Figure 12.37 *Effect of coolant flow on condenser duty*

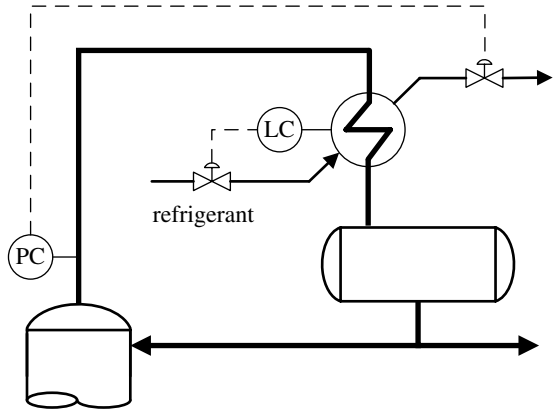

Figure 12.38 *Manipulating refrigerant flow*

shell side of the condenser is partially filled with boiling liquid refrigerant, thus removing heat as its heat of vaporisation. The column pressure controller manipulates the flow of refrigerant vapour returned to the compressor. An alternative configuration is to have the level controller manipulate this flow and the pressure controller manipulate the flow of liquid refrigerant. With either configuration there will be interaction between the controllers – usually resolved by tuning one controller a little slower than the other.

The system can become quite complex if designed to recover energy transferred to the refrigerant. The refrigerant vapour may be used to drive a turbine to provide pumping or compression elsewhere in the process. Or the refrigerant compressor discharge may be cooled by providing reboil energy for the column. As a result, disturbances can be propagated through the refrigerant system back to the column – requiring greater attention to the control design.

It is common for air to be the condenser coolant. Air-fin condensers comprise layers of finned tubes over which air is forced by fans. There are a number of ways in which air flow may be manipulated. As shown in Figure 12.39, the pitch of the fan blades or the fan speed can be adjusted. However, there are

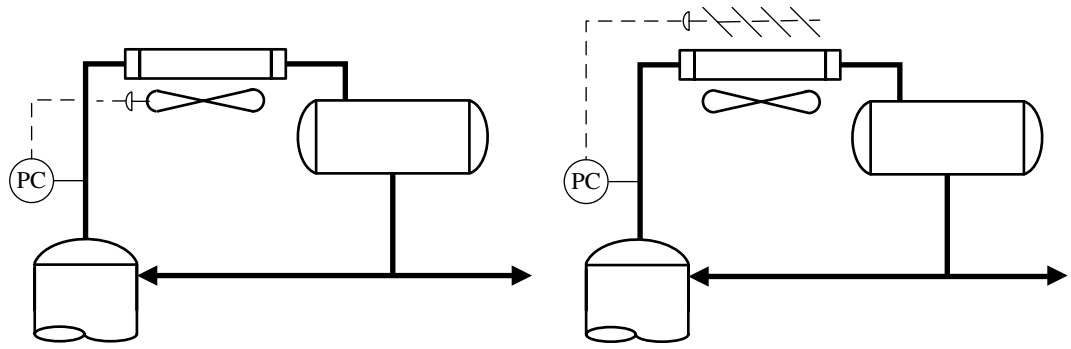

Figure 12.39 *Manipulating air flow*

usually a large number of fans. While it may not be necessary to install such mechanisms on all the fans, it does become costly and, like most mechanical systems, subject to reliability problems. It is also possible to change the number of fans in service. While this can be automated, it must be combined with some other more continuous variable to handle the situation where n fans are insufficient and $n + 1$ fans are too many. Another approach is to manipulate louvres placed in the stream of air exiting the condenser. This too is costly and subject to mechanical problems.

Air-fin condensers can be subject to rapid changes in ambient conditions. A sharp drop in temperature or a rainstorm can cause the reflux to be sub-cooled. On entering the column the reflux is brought back to its bubble point by condensing some of the rising vapour. This condensed vapour provides additional reflux; thus the *internal reflux* will be greater than the measured external flow. This then disturbs the product compositions.

It is possible to install an internal reflux controller. If ΔT is the temperature drop across the condenser, c_p the specific heat of the reflux material and λ its latent heat of vaporisation then

$$R_{internal} = R_{external}\left[1 + \frac{c_p \Delta T}{\lambda}\right] \tag{12.58}$$

As shown in Figure 12.40, we can modify the SP of the external reflux flow controller to take account of the sub-cooling and maintain a constant internal reflux flow.

While in principle this scheme appears to be beneficial, it may not perform well. One issue is timing. The reflux drum introduces a large process lag and so its exit temperature will change later than its inlet. Using the condenser outlet temperature would result in the reflux being corrected too early – although it would be possible to lag the measurement of temperature difference. Alternatively the drum exit temperature could be retained and the inlet temperature lagged.

The temperature difference can vary for reasons other than sub-cooling. For example an increase in the heavy key component in the overhead vapour will cause an increase in vapour temperature. The internal reflux controller will then reduce the reflux flow – the opposite of what is required to deal with the composition change. Using, instead of the overhead temperature, a constant set at a typical value can resolve this. But then the correction for sub-cooling, although directionally correct, will not be of the correct magnitude.

Moving to the next sub-category of schemes, another approach is to manipulate coolant temperature. While not applicable to air-fins it does offer some advantage if condenser fouling is an issue, since it

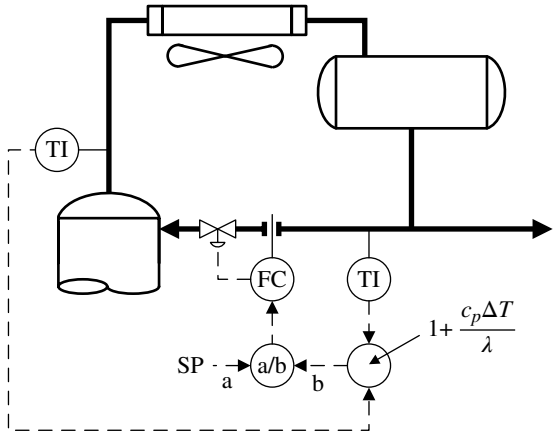

Figure 12.40 Internal reflux control

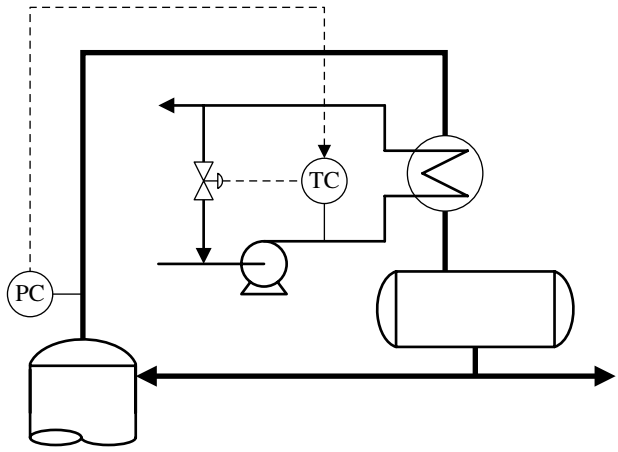

Figure 12.41 Manipulation of coolant temperature

permits a high coolant flow to be maintained – no matter what condenser duty is required. As can be seen in Figure 12.41, it does require additional pumping and so is more costly to implement.

The next sub-group of strategies are techniques which adjust the efficiency of the condenser. The first of these, shown in Figure 12.42, places a valve in the vapour line before the condenser. The pressure drop across the valve lowers the temperature at which the vapour condenses, hence lowering the LMTD across the condenser and therefore the heat it removes. It is a more costly approach since the valve is placed in what is probably the largest diameter line. Further, because the column and drum can now be isolated from each other, relief valves must be installed on both vessels.

Another way of reducing the efficiency of the condenser is to reduce the effective surface area used for heat transfer. Figure 12.43 shows how, by placing the control valve under the condenser, liquid can accumulate in the condenser. This *flooded condenser* is less efficient because less heat transfer takes place in the submerged part of the tube bundle. Here sensible heat is removed in sub-cooling the liquid. In the

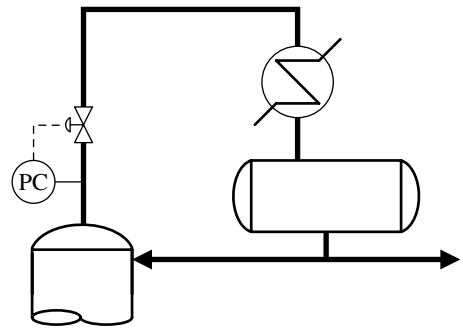

Figure 12.42 Reduction of condensation temperature

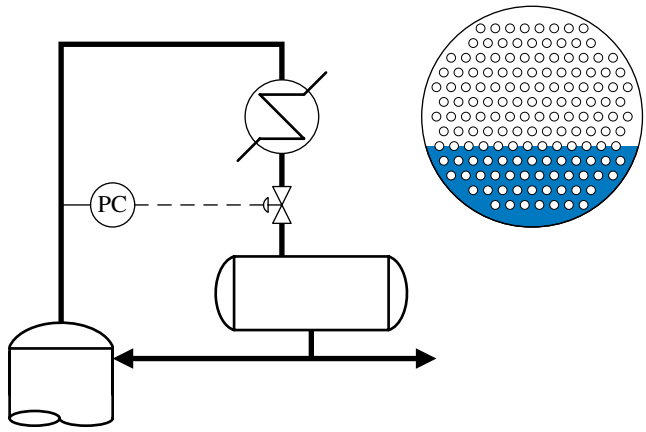

Figure 12.43 Flooded condenser

exposed part of the bundle, heat of vaporisation is removed condensing the vapour and the resulting condensate quickly leaves the tube surface. The pressure controller indirectly changes the level of liquid in the condenser.

While the flooded condenser approach offers little advantage if the coolant is liquid, it is beneficial if applied to air-fin condensers. It can replace the potentially unreliable mechanisms for manipulating air flow and is considerably less costly to install. With air-fin condensers it is the tube side that is flooded. It is better applied to single pass condensers. It is a myth that the condenser should be inclined to avoid the step changes in surface area that might occur as layers of tubes are exposed. It is unlikely that in the presence of the turbulence of condensation, and of the gradient of the liquid level necessary for flow to occur, that a full layer of tubes would be exposed before the liquid begins to drain from the layer below. Plus the condenser is likely to be inclined slightly in any case to ensure it is self-draining. The scheme is shown in Figure 12.44.

Conversion of any existing scheme to a flooded condenser is more than just a change in control configuration. It requires a full process design check and is likely to result in changes to relief valves and other safety-related systems. There are also a number of ways in which it can be configured. For example the

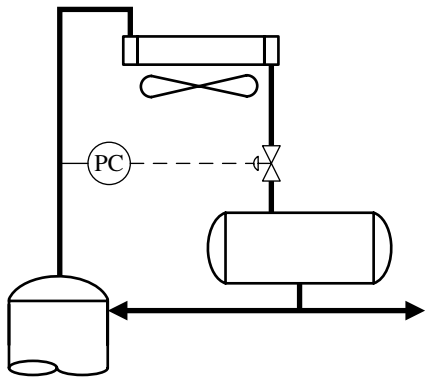

Figure 12.44 Flooded air-fin condenser

drum can also be flooded thus avoiding the need to install a valve on its inlet. The pressure controller can then manipulate either the reflux flow or the distillate flow directly.

If the drum is not flooded, then it is better if the condensate enters the drum above the liquid level to avoid interaction between the drum level controller and the column pressure controller. Otherwise a change in liquid level will affect the column pressure. A pressure equalising line between the column and the drum, although not essential, will keep the drum pressure constant. This is particularly beneficial if a vapour product is taken from the drum, since pressure variations would disturb the composition of both this and the liquid distillate product.

The drum can also be omitted. This gives a considerable cost saving on new plant. It also eliminates a major source of process lag thus allowing the composition controllers to be tuned to act much more quickly. Its disadvantage is that, in the event of a process upset, there is very little liquid inventory.

The two possible drum-less schemes are shown as Figure 12.45. The decision as to whether to control the liquid level (in the condenser) by manipulating reflux or distillate flow is part of a much more wide-ranging consideration that will be covered in the next section. However, controlling pressure by manipulating the reflux flow helps considerably with disturbances to the energy balance, such as those caused by changes in ambient conditions around the air-fin condenser. We saw earlier in this chapter that distillate cut should be kept constant during such disturbances. If for example there is a rainstorm, the condenser duty will increase – condensing more vapour and reducing column pressure. The pressure controller will

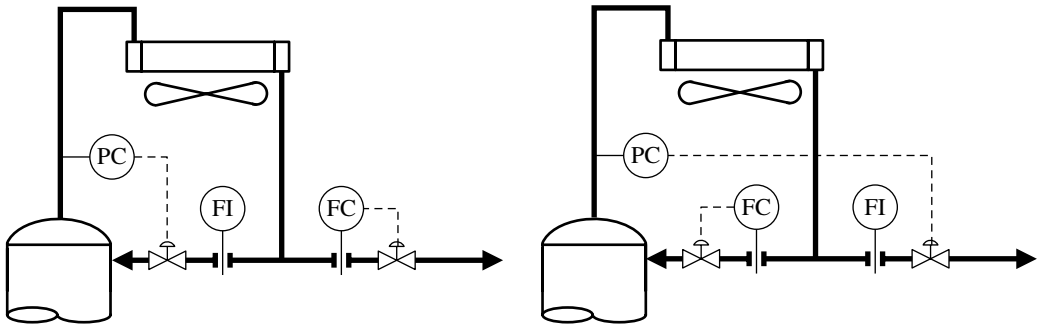

Figure 12.45 Drum-less flooded condenser

resolve this relatively quickly by reducing the reflux to build level in the condenser. In doing so it helps compensate for the drop in reflux temperature and so maintain a more constant internal reflux. With the cut kept constant by the distillate flow controller, and the fractionation kept approximately constant by the reduction in external reflux, both product compositions will remain relatively unchanged. However, as we shall see later, there are several other factors to consider before selecting this configuration.

Figure 12.46 shows an alternative configuration for a flooded condenser. The control valve is located before the condenser. It is important that the condensate enters the drum below the liquid level and the equalising line between column and drum is essential. If the pressure in the column rises above SP the pressure controller will open the valve. This increases the pressure in the condenser. Since the pressure in the drum is unaffected the liquid level in the condenser falls until the head of liquid matches the pressure drop across the control valve. This exposes more tubes and therefore brings down the pressure.

From an operator perspective the scheme is conceptually more difficult to understand. It requires a larger, and hence more costly, control valve because it is located in the large diameter vapour line rather than the smaller diameter condensate line. The process design is also more difficult. Since it offers nothing over the alternatives it should not be considered further.

Our last sub-group of techniques, which adjust the quantity of vapour condensed in the condenser, is known as *hot gas bypass* or *hot vapour bypass*. It can be implemented in several ways; the first of which, shown in Figure 12.47, we have already discussed as an option with the basic flooded condenser approach. It includes the addition of what we previously described as an equalising line, but can also be thought of as a vapour bypass. It may not be immediately obvious why bypassing the condenser affects pressure since all the vapour product is condensed no matter how far the bypass is opened. However, since no change is made to the coolant flow, routing less vapour to the condenser will result in increased sub-cooling of the condensed liquid. This causes the temperature in the reflux drum to fall and the bypassed vapour now condenses in the drum. Because of the resulting much reduced efficiency in heat transfer the pressure will increase to a new equilibrium where the resulting increase in dew point, which makes the vapour easier to condense, is sufficient to compensate for the reduced efficiency. Care is needed in the process design to ensure good mixing between the bypassed vapour and the sub-cooled condensate. Otherwise the reflux temperature is likely to fluctuate, causing disturbances to the internal reflux.

Figure 12.48 shows the same scheme but with the condenser located below the drum. This has advantages in terms of easier access for maintenance work. Because it is the smaller line it is tempting to reduce

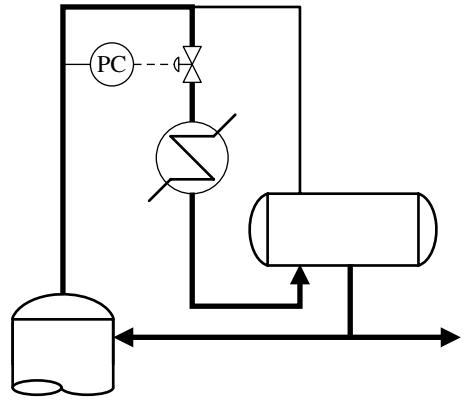

Figure 12.46 *Alternative configuration for flooded condenser*

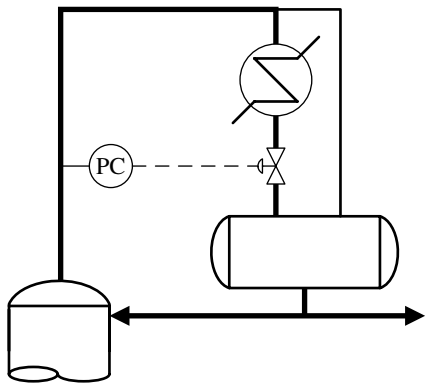

Figure 12.47 Hot gas bypass

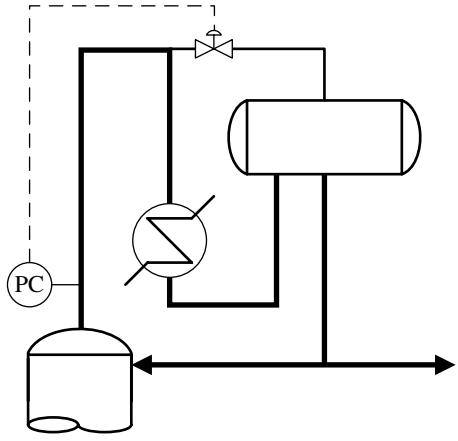

Figure 12.48 Hot gas bypass with condenser below drum

the installation cost by locating the control valve in the bypass rather than in the condensate line. This, however, can cause problems with inverse response. This is illustrated in Figure 12.49.

When the valve is opened two competing processes take place. The first is one of material transfer from the column to the drum. This has the effect of reducing column pressure. The second, because of the bypass being opened, is one of heat transfer reducing the amount of vapour condensed and so increasing pressure. Because the dynamics of material transfer are generally faster than those of heat transfer we see the first of these effects. However, the drum pressure rises quickly and the material transfer slows. The heat transfer process ultimately prevails and the pressure rises above that at which it was before the control valve moved. The amount of inverse behaviour depends on the relative dynamics of the two processes. On some columns it may not be noticeable; on others it may be severe.

While it is possible to tune a PID controller in this situation, it is necessary to greatly reduce the controller gain – thus making it very slow to respond to disturbances. An alternative solution is to relocate the pressure transmitter from the column to the drum, as shown in Figure 12.50. While the two processes affecting pressure are unchanged, they no longer compete. While a simple, low cost solution – particularly

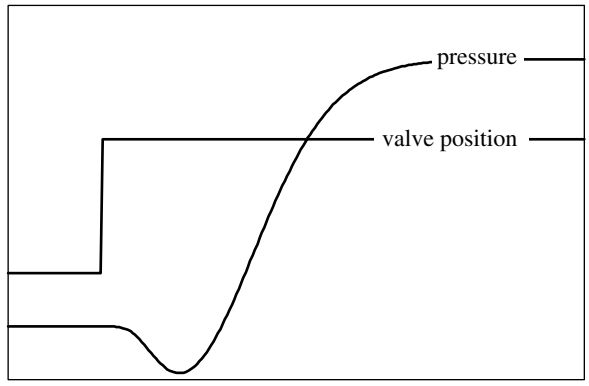

Figure 12.49 Inverse response

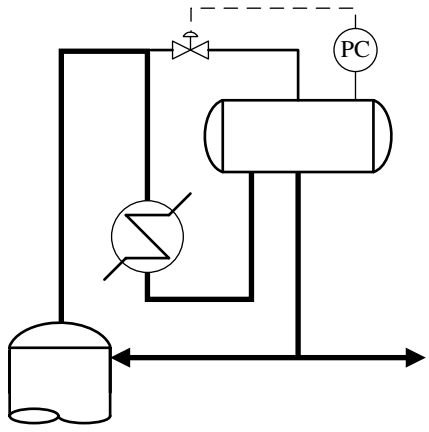

Figure 12.50 Relocating pressure transmitter

if trying to resolve a problem on an existing unit, its limitation is that it does not control the pressure in the column. While the variation is likely to be relatively minor, it will have an effect on product composition.

The ideal solution is to locate control valves in both locations, as shown in Figure 12.51. That in the condensate line is used to control column pressure, while that in the bypass used to control drum pressure. If the condenser is below the drum there must be sufficient difference between the two pressures to overcome the maximum liquid head. If this is not the case then, on high pressure, the column pressure controller will saturate and the pressure will rise until it is sufficient to overcome the head. While not necessarily unsafe, it does mean that full control of pressure will be lost.

Rather than rely on the process operator to maintain sufficient pressure difference the drum pressure controller may be replaced by a differential pressure controller (dPC), as shown in Figure 12.52. The SP of this controller could be fixed and not adjustable by the operator, or a safe minimum limit configured. It might also include logic that disables column pressure control if the dPC is switched to manual.

The two controllers will interact. If column pressure rises above its SP the column pressure controller will open the condensate valve so that more vapour passes through the condenser. But, as the column pressure rises, the measurement of the dPC will also increase and its controller will respond by opening the

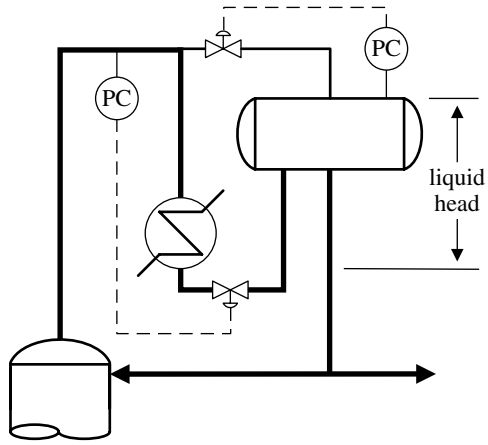

Figure 12.51 Use of two pressure controllers

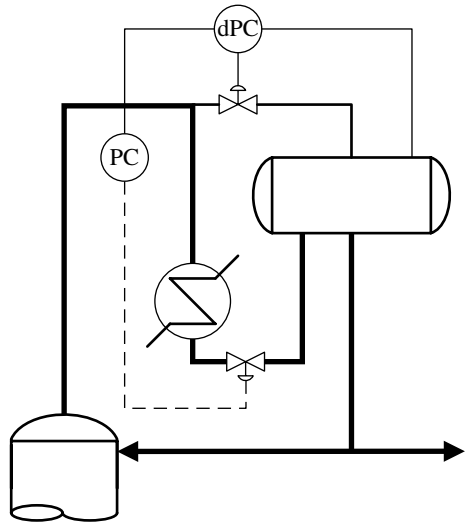

Figure 12.52 Pressure difference controller

bypass. This will reduce the flow of vapour to the condenser – the opposite of what is needed. To help break the interaction the dPC should be configured to use the SP of the column pressure controller, not its PV.

The next main group of schemes manipulate the vapour flow leaving the process. Clearly these only work if the condenser is not a total condenser. But, if there is a significant flow of vapour, adjusting it has the most direct control of pressure and will have an immediate effect. How we actually manipulate the flow will depend on the vapour handling system. If the vapour is simply routed to a lower pressure system then we need only place a control valve in its line, as shown in Figure 12.53 – for both noncondensing and partial condensing situations.

If the overhead vapour is routed to a compressor then all of the schemes described in Chapter 11 can be applied. These include suction or discharge throttling, manipulating speed or inlet guide vanes and the use

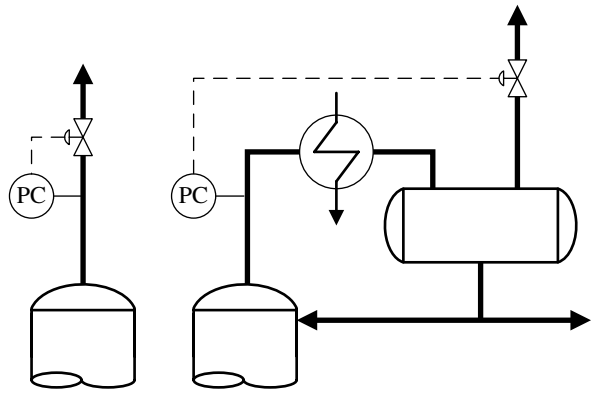

Figure 12.53 *Manipulation of vapour flow*

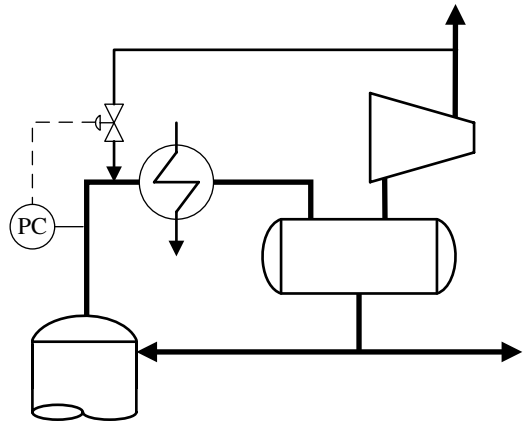

Figure 12.54 *Manipulation of compressor spillback*

of recycle (or *spillback*). In the latter case the spillback can be cooled through a dedicated exchanger or routed back to the condenser inlet, as shown in Figure 12.54.

The pressure in vacuum distillation columns is similarly controlled by manipulating the spillback around the ejectors, as shown in Figure 12.55. The installation of the pressure transmitter needs special attention. It is important that the impulse line is self-draining back to source; otherwise a liquid head can build up and cause a false pressure measurement. The liquid may also boil and cause a noisy measurement.

The ejector will become unstable if the flow of motive fluid is too low and *choking* will occur if too high. For these reasons it is common to keep this flow fixed. Because ejectors operate well only at design pressure drop, similar problems can arise with throttling either the inlet or discharge. This leaves spillback as the only viable manipulated variable. However, it is common for this to be closed and the pressure controller to be on manual. This is not a reflection on the performance of the scheme. It can be economically very attractive to operate the column at the lowest possible pressure, even if this means the pressure fluctuating somewhat. We will cover later in this chapter techniques for compensating for such fluctuations so that product composition is not affected. And we will also return later to pressure optimisation.

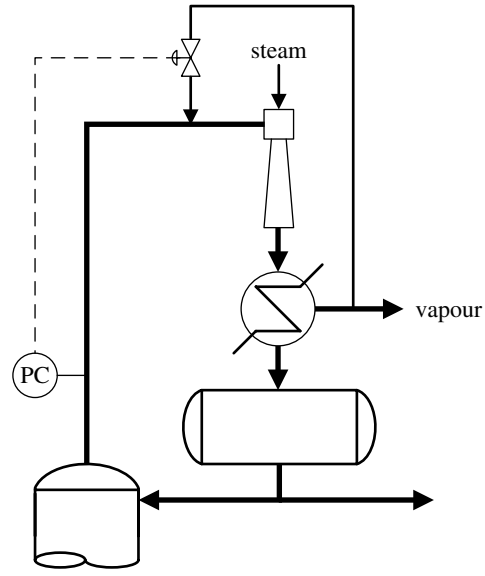

Figure 12.55 Manipulation of ejector spillback

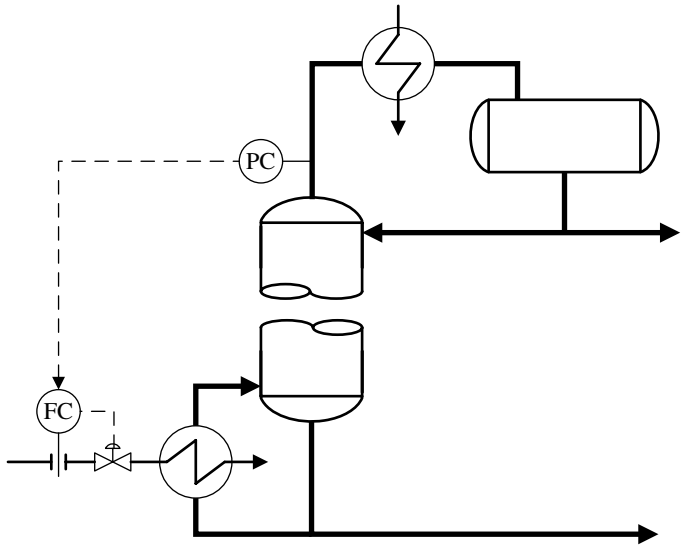

Figure 12.56 Manipulation of vapour generation

The last of the strategies we might consider for pressure control is manipulation of the rate at which vapour is generated. A typical scheme is shown in Figure 12.56. This scheme should only be applied under special circumstances. While quite feasible it does result in the loss of the main fractionation variable as a means of controlling product composition. But there are columns where tight control of composition is not required – maybe because they are deliberately over-fractionated.

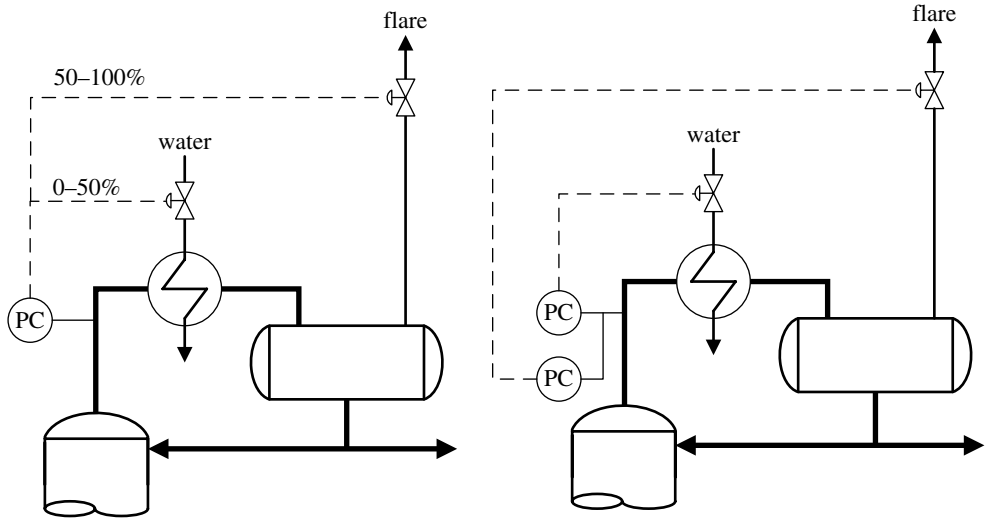

Figure 12.57 Preferred alternative to split-ranging

The use of split range pressure controllers is common on distillation columns. They enable the operating range to be extended beyond the point where a MV saturates. One example, covered in Chapter 5, dealt with controlling pressure by manipulating vapour production, but also importing a non-condensable stream when completely closing the vapour off-take valve was not enough to raise the pressure to its SP.

Another example would be venting vapour if the condenser limit has been reached. In Chapter 5 we explained why it was often better to install two independent controllers, rather than the split range approach. Figure 12.57 shows how this would work on our case study column. The split range option, on rising pressure, is configured to first open the cooling water valve. If the controller output reaches 50% the cooling water valve will be fully open and, if necessary, the valve to flare will begin to open. The preferred design has a controller manipulating the cooling water valve with a SP slightly lower than that manipulating the flare valve. As described in Chapter 5, it is important the two controllers use the same pressure transmitter.

12.8 Level Control

After maintaining the energy balance, the next prime objective of the basic column controllers is to maintain the material balance. Material enters as feed and leaves as products. If these are not in balance, then the inventory in the process will change. This is reflected by changing levels in the reflux drum and/or the column base. By controlling these levels we maintain the material balance across the column.

So far we have used up one of our available MVs (condenser duty) to control one of our PVs (pressure). There remain four MVs from which we can select two to provide control of our two levels. To these we can add feed rate, since on some columns this is available as an MV. Table 12.6 shows that there are in theory 20 potential schemes.

Two of the potential schemes can be rejected immediately, on the grounds that they do not meet the objective of maintaining the material balance across the column. Figure 12.58 shows one of these (see note 1 in Table 12.6). A change in feed rate will cause the column to move out of material balance but, since both product flows are fixed by flow controllers, no corrective action can be taken. For example, an

Table 12.6 *Potential level control strategies*

		Secondary of column level controller				
		Feed	Distillate	Reflux	Bottoms	Reboil
Secondary of reflux drum level controller	Feed	×	Impractical	Impractical	Impractical	Impractical
	Distillate	Not always available	×	Impractical	Energy balance scheme (4)	Material balance scheme (5)
	Reflux	Not always available (3)	Impractical	×	Material balance scheme (6)	Violates material balance (1)
	Bottoms	Impractical	Impractical (2)	Impractical	×	Impractical
	Reboil	Impractical	Impractical	Violates material balance	Impractical	×

increase in feed rate will cause the column level to increase. The column level controller will increase the reboiler duty, which will increase column pressure. The pressure controller will then increase condenser duty. The additional condensate will cause the reflux drum level to rise and its level controller will return it to the column, where it will again cause the column level to rise. If left unchecked the column will ultimately fill with liquid. The second scheme rejected again has both product flows fixed – compounded by a rather strange configuration of level controllers.

Many of the schemes are described as 'impractical'. One of them (note 2) is shown as Figure 12.59. Many engineers would reject the scheme instinctively; it simply does not look 'right'. However, it would function correctly provided one of the level controllers is configured as reverse acting, i.e. on increasing PV it will reduce its output. In this example the column level controller is reverse-acting. So, if the operator were to reduce its SP, it will reduce the distillate flow. The drum level will then begin to rise and so its controller will increase the bottoms flow – bringing down the column level as required. Similarly, if the operator were to reduce the SP of the drum level controller, it would increase the bottoms flow. The column level will fall causing the level controller to increase the distillate flow and so bring down the drum level. What makes the scheme impractical is that each level controller relies on the other to take corrective action. If one of the controllers were switched to manual, then the output of the other would eventually saturate. While in some instances it is unavoidable, it is not generally advisable to design such *nested* controllers – particularly mutually nested controllers. All of the schemes described as impractical in Table 12.6 involve unnecessary nesting.

Two of the schemes involve manipulation of the feed rate. On many columns this is not an option since they receive their feed from a process upstream that has its feed rate fixed to meet some other criterion. Even on those which receive feed from tankage, manipulating feed rate may be undesirable. Operating plans will often specify feed rather than product rate. Nevertheless there are occasions where the scheme is used. There are examples of a series of several columns all having level controllers cascaded to each of the feed flow controllers. Figure 12.60 shows one of the two possible schemes (note 3). It is often used when both products are routed to downstream processes and it is a requirement that the feed to the processes is kept constant. In reality this is probably not achievable. While we can manipulate reboiler duty to provide control of composition, this will only vary separation. To vary the cut at least one of the product

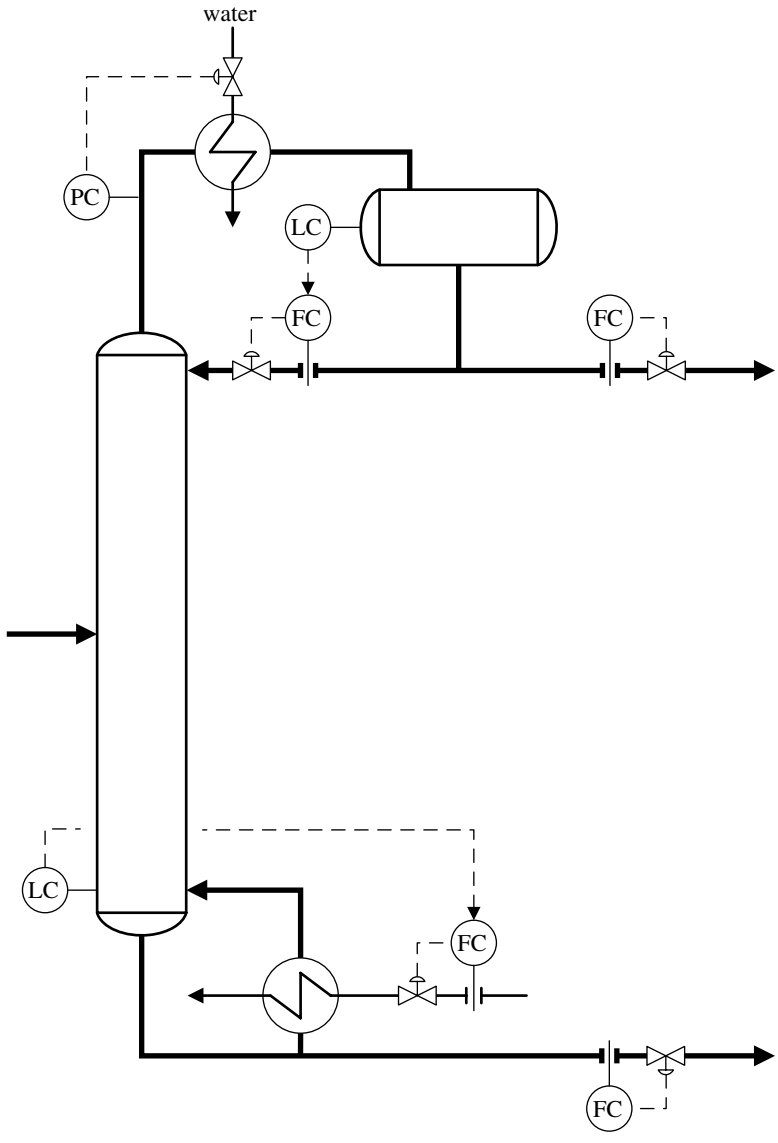

Figure 12.58 *Violating the material balance*

flows will need adjustment. One could argue that it would be better to cascade the drum level to the distillate flow. This would permit averaging level control to be applied to make smooth changes to the downstream process flows and release the reflux flow as a MV for composition control.

Of the three schemes remaining one (note 4) has both product flows as MVs of level controllers, as shown in Figure 12.61. This is known as the *energy balance* scheme. The remaining MVs, reboil and reflux, will then be used for composition control. Both MVs affect both cut and separation. So, unlike other configurations, it is not simple to vary cut without changing separation or vice versa. This does not imply that there is any problem with the scheme; indeed, it is the most commonly used configuration.

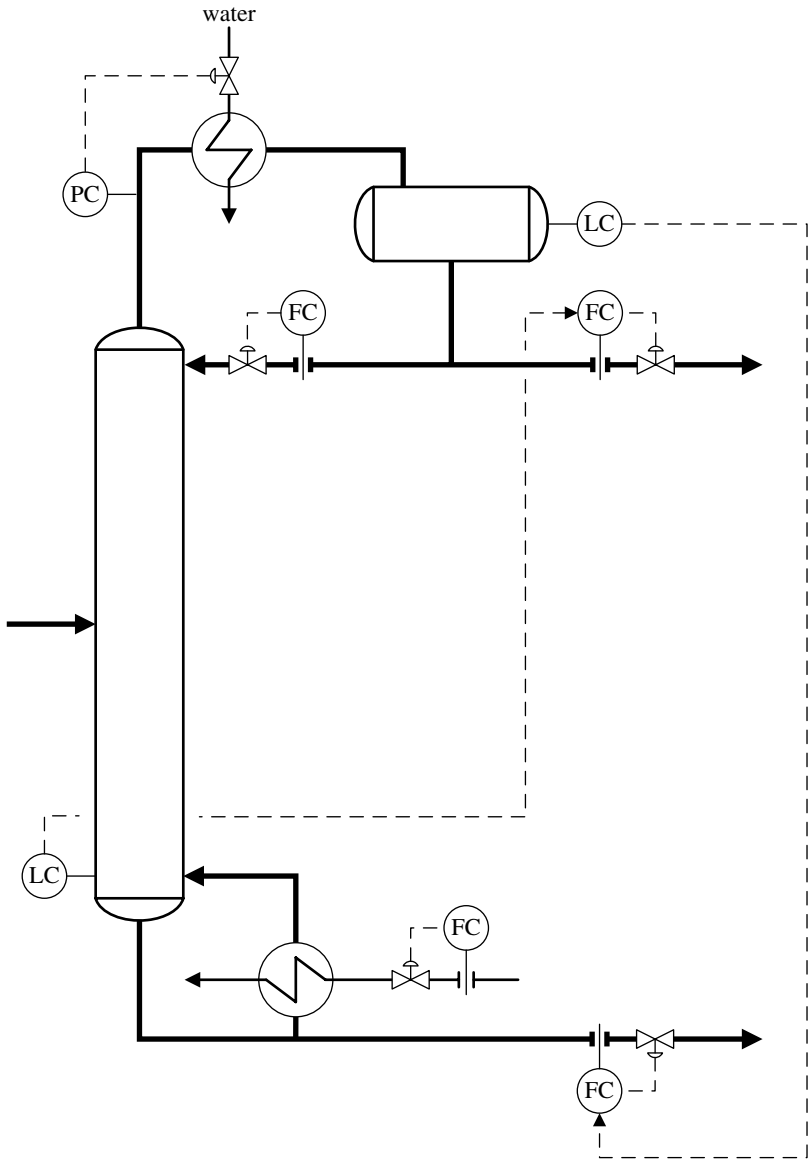

Figure 12.59 *Impractical level control configuration*

The two remaining schemes are known as *material balance* schemes. One of the product flows is not used for level control and remains available as a MV for composition control. The first of these schemes (note 5) is shown in Figure 12.62. This scheme is undesirable for a number of reasons. Firstly, it can exhibit inverse response; additional reboiler duty causes the vapour volume in the reboiler to increase and so displace some liquid into the column base. This can also be aggravated by the additional vapour displacing liquid on the lower trays. The level may then initially rise before falling as a result of the increased vaporisation. Less severe inverse behaviour may not be apparent but will still cause an increase in deadtime.

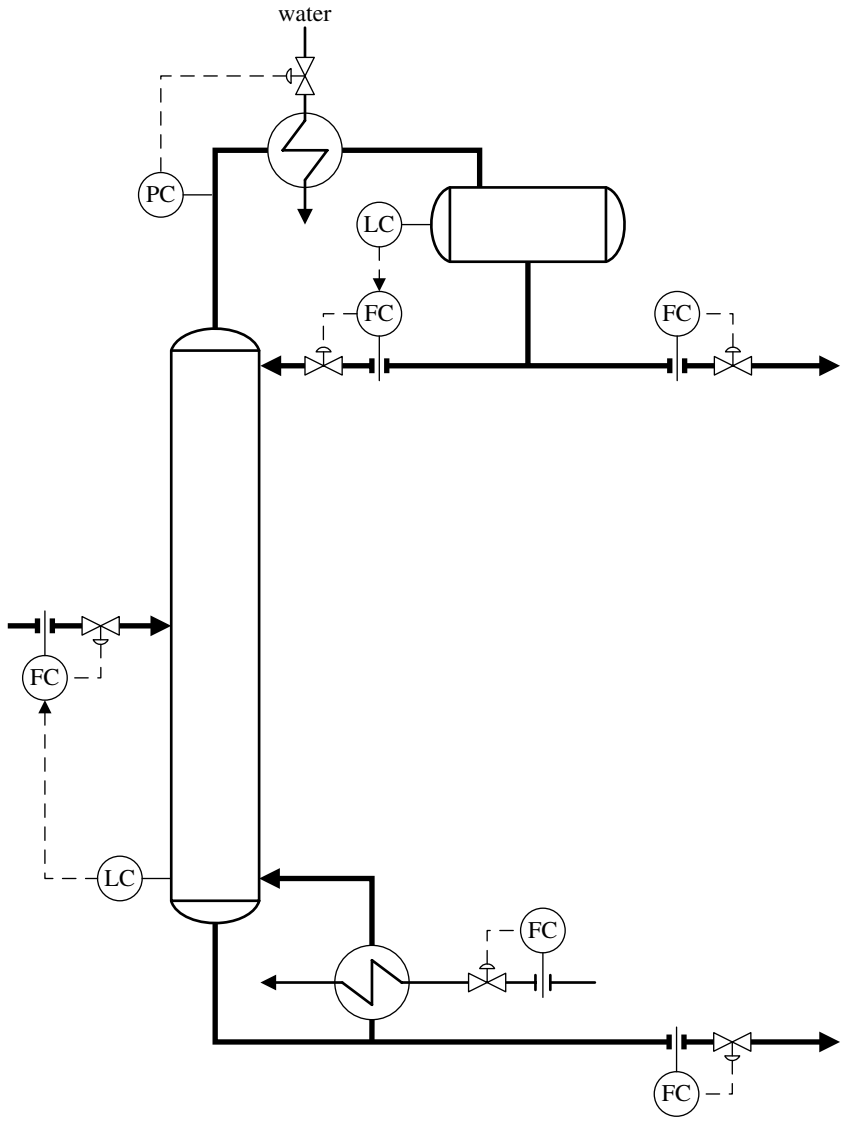

Figure 12.60 *Using feed rate to control level*

Secondly, the thermal inertia of the reboiler will introduce a large lag, not common in level control. Controller tuning must therefore be relatively slow. It may be that it cannot respond sufficiently quickly to routine changes in reflux or bottoms flow. So, to avoid violating level alarms, changes have to be made slowly – thus degrading the performance of the composition controllers that manipulate these variables. However, there are occasions where use of the scheme proves necessary. For example, if the bottoms flow represents only a small part of the feed, manipulating it over its full range may not be sufficient to control column level – leaving the reboiler as the only feasible MV. The MVs remaining available for composition control are bottoms flow, which determines cut, and reflux flow, which determines separation.

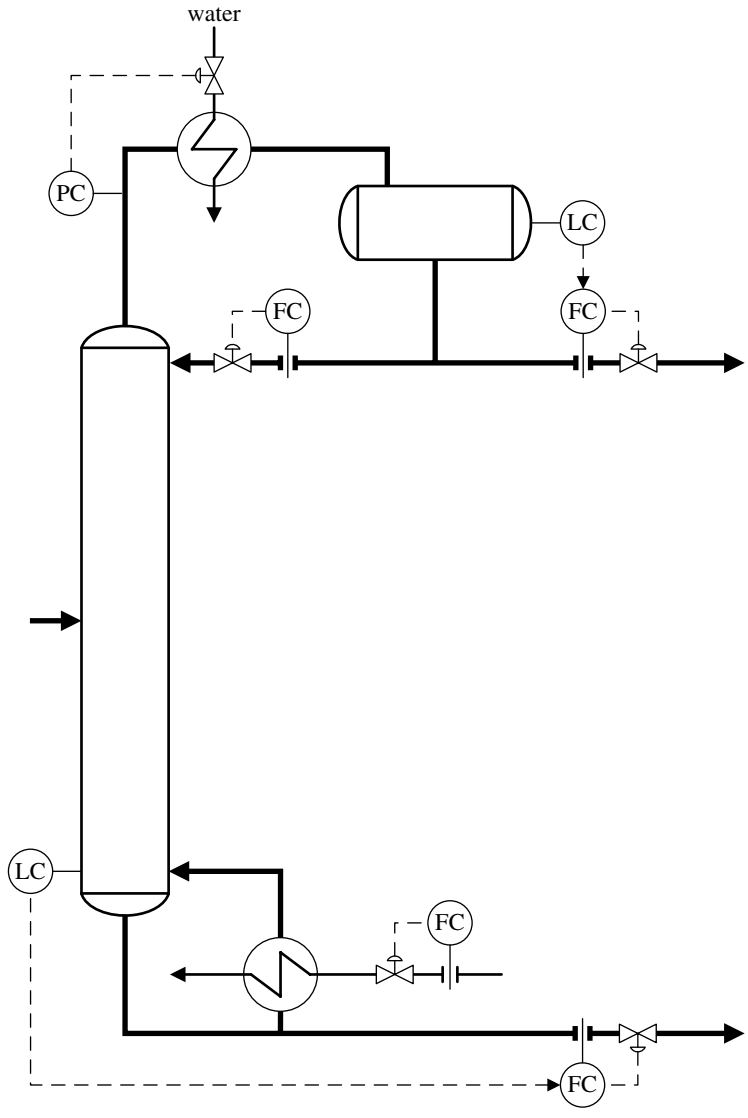

Figure 12.61 *Energy balance scheme*

Figure 12.63 shows the more common version of the material balance scheme (note 6). This does not share the problems with the previous scheme and thus is in common use. The remaining MVs that will be used for composition control are distillate flow, which determines cut, and reboil duty, which determines separation.

The naming of the 'energy' and 'material' balance schemes can cause confusion. Any scheme must ensure that both energy and material balance in terms of what enters and leaves the column. The naming arises from the variables that remain to control product composition. In the material balance schemes, the material balance is one of the variables manipulated directly (by changing distillate or bottoms flow).

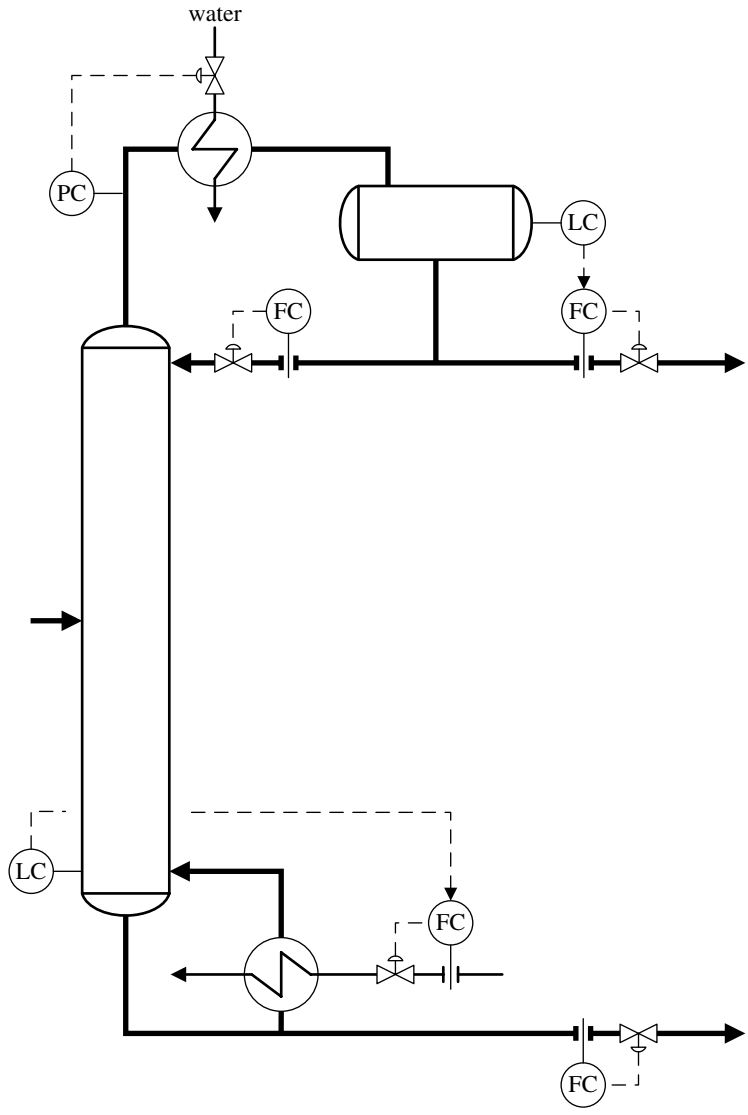

Figure 12.62 *Material balance scheme (less preferred)*

In the energy balance scheme, only the energy balance can be manipulated (because all the product flows are used for level control).

The result of the exercise just completed is that, in designing a level control strategy, the choice in the majority of cases is restricted to one of two options. Either we apply the energy balance scheme and cascade the drum level controller to the distillate flow, or we select the preferred material balance scheme and cascade the level controller to the reflux. There are a number of considerations in making this decision.

We have seen that cut is the prime variable in determining product composition. With the material balance scheme in place, cut can only be varied by changing the distillate flow. However, this in itself has no effect on product composition. It merely changes the flow of material that has already been produced

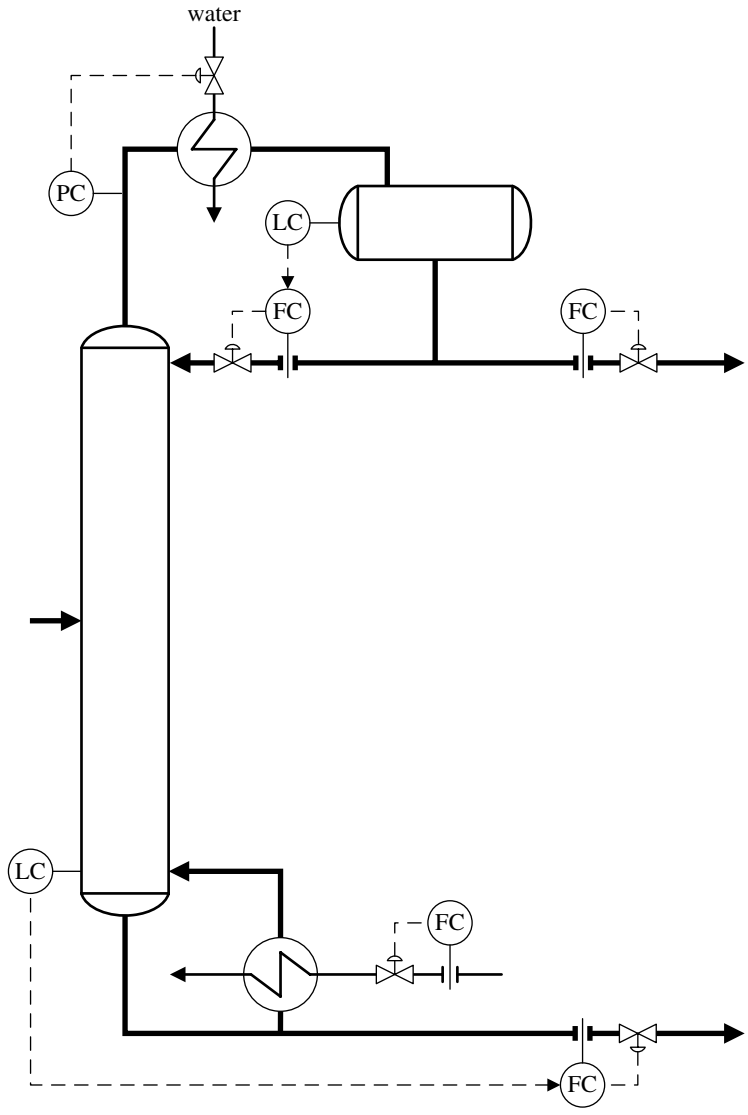

Figure 12.63 *Material balance scheme (preferred)*

in the reflux drum. But changing the distillate flow causes the drum level to deviate from SP and it is the corrective action of this controller changing reflux that changes the composition. To achieve tight control of composition we therefore need tight level control. In the case of the energy balance scheme, we are free to choose tight or averaging level control, since the composition controller will manipulate reflux directly. If the distillate product is routed to a downstream process, averaging level control will exploit the surge capacity of the reflux drum to minimise flow disturbances. Even if the distillate product is routed to tankage, if it is cooled via heat integration, averaging level control will minimise disturbances to the energy balance.

A similar argument can be applied to the choice of energy balance over the less preferred version of the material balance scheme. In theory it allows us to install averaging tuning in the column level controller. However, if the holdup in the column is small, this will give little advantage. Further, depending on the reboiler configuration, permitting the level to vary may cause problems with its performance. Nevertheless there are occasions where averaging level control will be beneficial.

Another consideration in making the choice is the reflux ratio (defined in this book as R/D, not $R/(R + D)$ as it is in some). A large reflux ratio favours the use of the material balance scheme because the relatively small distillate flow may not offer the range necessary to adequately control drum level. Similarly a low reflux ratio ($<< 1$) favours the use of the energy balance scheme.

The material balance scheme is favoured when it is important to maintain the cut constant. Both versions of the scheme do this. We have seen that this is desirable when the prime disturbances arise from disruption to the energy balance – such as rainstorms on air-fin condensers, changes to feed enthalpy, changes in column pressure and variation in reboiler duty due to heat integration.

We have seen that we can define cut both in terms of product yield but also in terms of temperature. A tray temperature controller can be configured to manipulate whatever variable remains after the level controllers are configured. And so we can control cut whether we have selected either the energy balance or the material balance scheme. However, there are occasions when tray temperature is insensitive to changes in product composition – for example when separating components with very similar bubble points. Under these circumstances the cut control provided by the material balance scheme is advantageous.

The energy balance scheme is favoured when the main disturbances arise from changes to the material balance. For example, a change in feed composition requires that the cut be changed in order to maintain product compositions on target. The energy balance scheme permits the cut to change. While it is unlikely to change the cut to the correct value it does change it in the right direction and will outperform either material balance scheme. For the same reason, the energy balance scheme will perform better during feed rate changes.

We have seen that the distillate composition controller on the preferred material balance scheme relies on the level controller. If this is switched to manual, then the composition control should be automatically disabled and not permitted to be re-commissioned until the level is back on automatic. Similarly, on the less preferred version of the material balance scheme, bottoms composition control should not be permitted if the column level controller is switched to manual. Without this precaution the composition controller will ramp its output until it saturates.

A limitation of the material balance control scheme is that if the product flow is not used as a MV for composition control, for example if the controller is switched to manual, it may not be possible to control the composition of the other product. The cut may be such that achieving the target composition is infeasible. This problem can also arise dynamically. If the composition controller manipulating cut is much slower than that manipulating separation, then the latter may be need to be tuned to act slowly enough for the other to first correct the cut. This is common if the distillate composition is controlled using an on-stream analyser to manipulate the distillate flow. The lag and deadtime imposed by the reflux drum and analyser mean that control will be slow compared to the other composition controller which might comprise an inferential manipulating reboil duty.

The material balance scheme is often favoured on high purity columns. We have seen that the permitted range of cut is much smaller under these circumstances. A scheme that keeps tight control of cut should therefore perform better. High purity columns also tend to have high reflux ratios, further favouring the use of the material balance scheme.

On many columns there can be conflicting arguments for the choice of scheme. Here the approach should be to make a preliminary selection of one of the schemes, identify its limitations and attempt to

enhance the scheme to deal with these. If this fails, then switch to the alternative and enhance this one. For example we might have good reasons to select the material balance scheme but the column is subject to changes in feed rate. Installation of the feedforward scheme shown in Figure 12.64 will maintain a constant *D/F* ratio and so overcomes this limitation. While not quite as simple as drawn, a full description of feedforward control is presented later in this chapter.

The same column might also be subject to changes in feed composition. If we are in a position to measure this on-stream, or infer it, then the proportion of light key in the feed (LK_f) can be fed forward to the ratio controller target by applying Equation (12.25) – using the target values of LK_b and LK_d. Figure 12.65 shows a simplified version of the scheme.

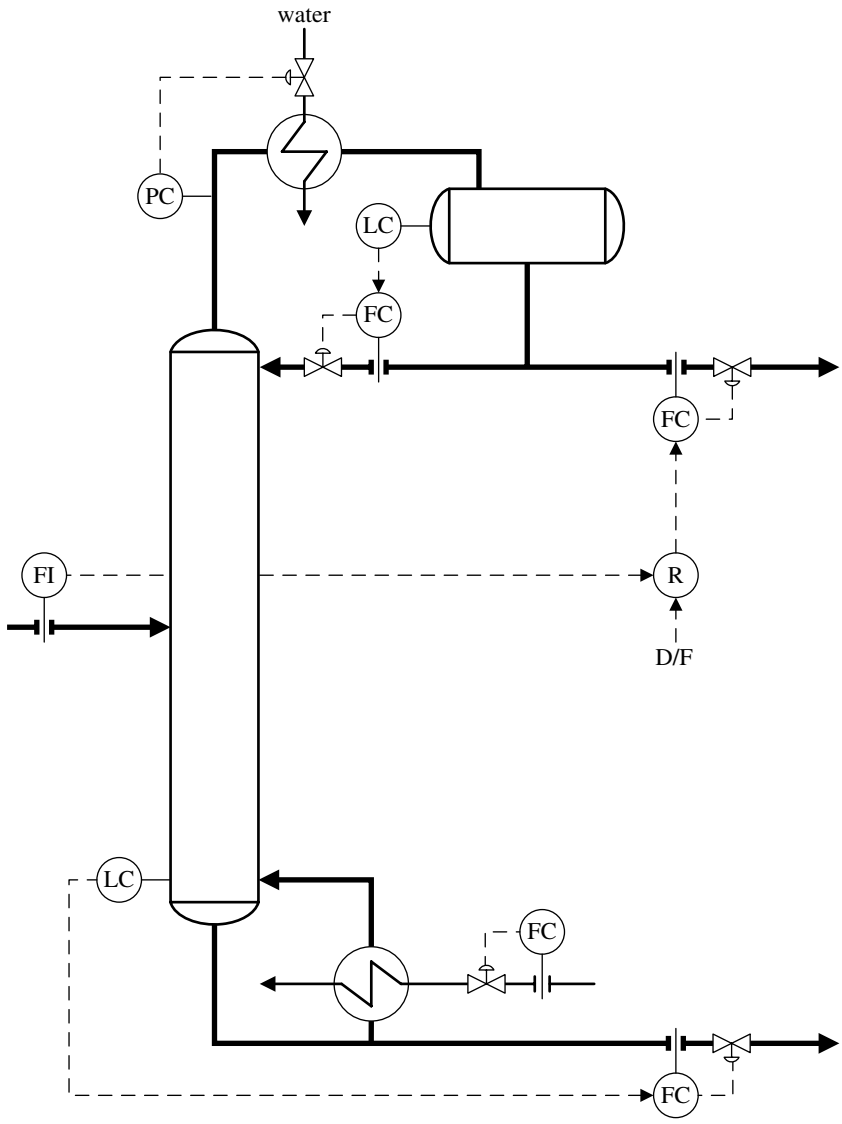

Figure 12.64 *Enhancing the material balance scheme to maintain cut*

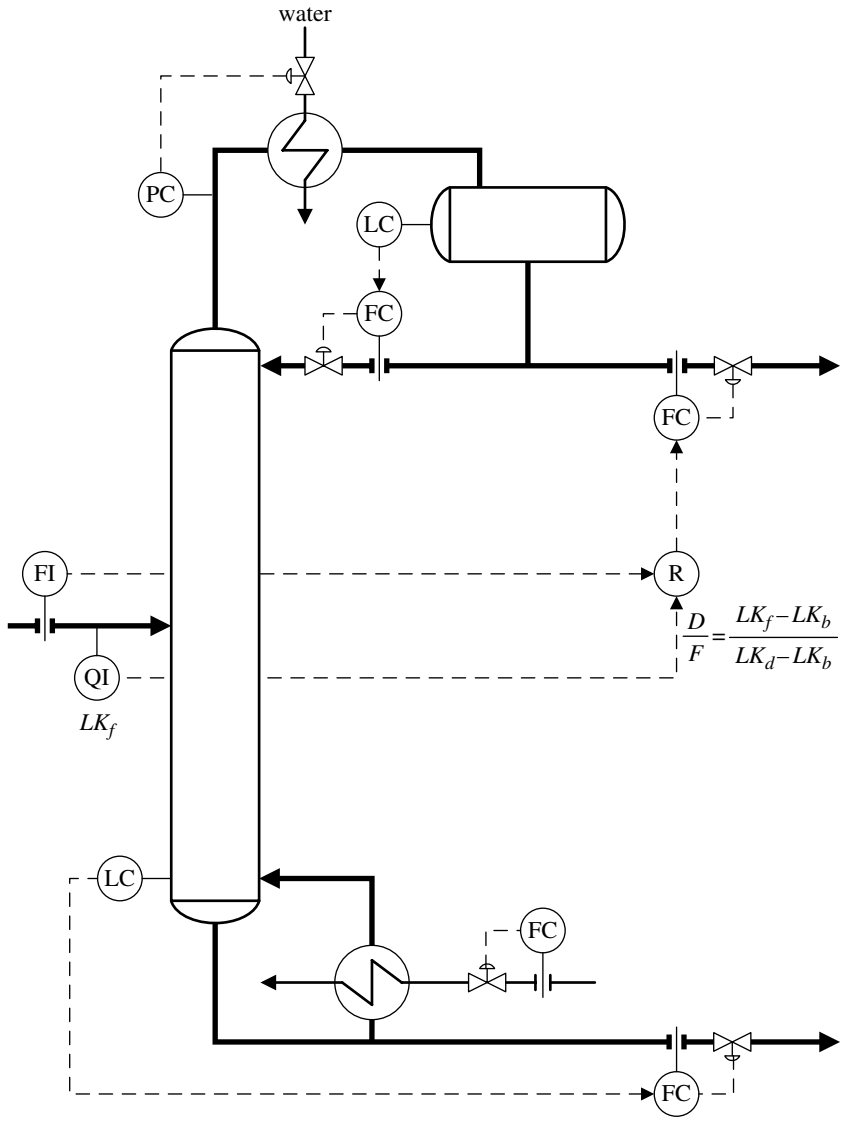

Figure 12.65 *Enhancing the material balance scheme to adjust cut*

Another modification can be made to the material balance scheme aimed at overcoming the lag introduced by the reflux drum. This is shown in Figure 12.66. Any changes to the SP of the distillate flow controller made to correct composition are passed directly to the reflux flow controller SP, rather than wait for the drum level controller. If the level controller is tuned tightly according to the method given in Chapter 4, this would also be achieved by a conventional level controller. However, the scheme includes an additional term (K). If K is set to -1, the change in distillate flow SP is passed to the reflux flow SP as an equal and opposite change. As a result there is no change in drum level. However, K can be set as required. For example, setting a value less than -1 (e.g. -2) will result in the reflux being changed by more

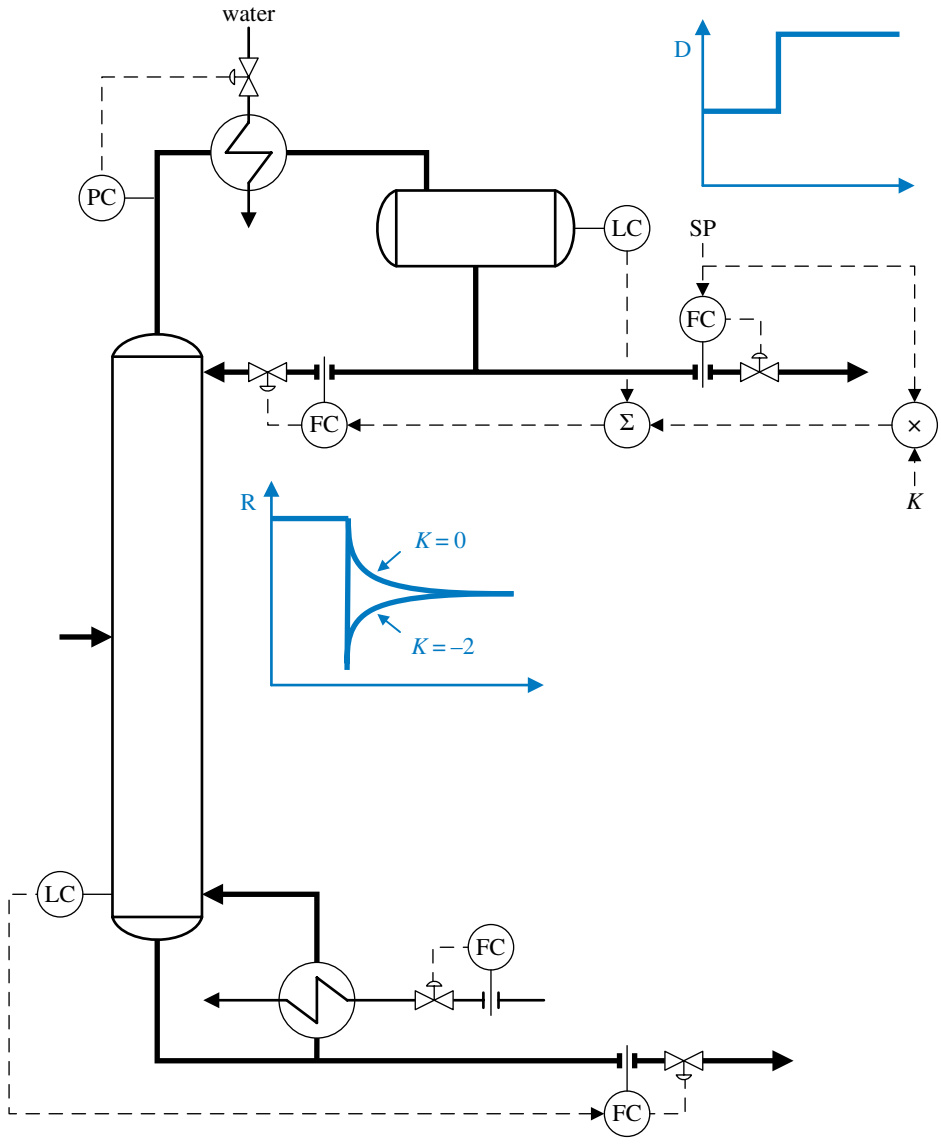

Figure 12.66 *Compensating for reflux drum lag*

than the change in distillate. The drum level will then change and the controller will take corrective action to bring the reflux back to the correct value. The 'kick' this introduces will help overcome the overall process lag. The scheme has introduced something similar to a lead-lag algorithm, as described in Chapter 6, where K is effectively the $T1/T2$ ratio. The lag is governed by the speed of response of the drum level controller. Tuning would be by trial-and-error, without an obvious measure of how well the scheme is performing. If such dynamic compensation is justified it would be more straightforward to use a conventional lead-lag algorithm and tune it according to the method given in Chapter 6.

The alternative approach of selecting the energy balance scheme and checking for its limitations might show that it does not handle well disturbances to condenser and reboiler duty. We have seen (Figure 12.40) how the use of internal reflux control might be applied to the first of these problems. If the disturbances to the reboiler duty arise from heat integration we can adopt a similar approach as shown in Figure 12.67. Thinking of the reboiler duty SP in energy units, we use the temperature drop across the reboiler to calculate $c_p \Delta T$ (where c_p is the specific heat of the heating fluid). Dividing the duty SP by this value generates the flow SP. Thus any disturbance to the temperature of the heating fluid will be immediately compensated for by the change in flow. While the lag of the reboiler means that outlet and inlet temperature do not change at the same time, the scheme will still work well. Lagging the inlet temperature to compensate for the reboiler lag would be counter-productive since it would delay the flow correction. An alternative approach is shown in Figure 12.68 where the duty is calculated and is used as the PV of a duty controller. It will perform in exactly the same way but, because it replaces rather than supplements, the flow controller it tends to be less popular.

Some sites have a hot oil system, where oil is heated using a fired heater and fed to a header running around the site. Several reboilers may then each withdraw oil from the supply header, returning it after use

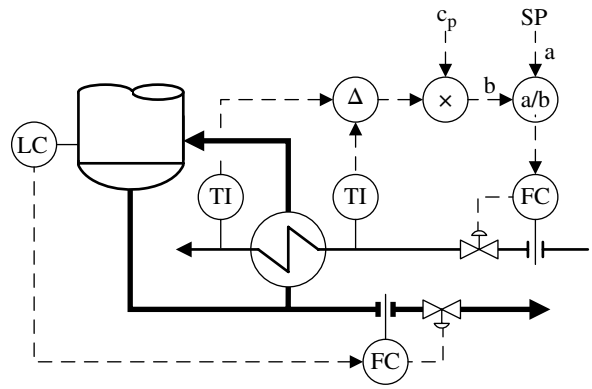

Figure 12.67 *Enhancing the energy balance scheme*

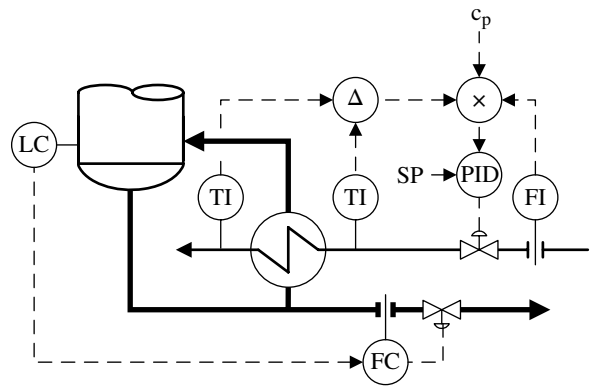

Figure 12.68 *Alternative reboiler duty controller*

to a separate return header. The oil supply temperature is set by a controller on the fired heater and so therefore not normally a source of disturbance. However there is often a minimum return temperature below which the increased viscosity impacts on the circulating pumps. The scheme in Figure 12.69 shows how this is managed. The flow may be fixed and reboiler duty adjusted by manipulating this temperature (subject to a minimum SP) – remembering that an increase in duty requires this temperature to be reduced. Alternatively the temperature may be fixed and the flow manipulated. Either way, some logic may be necessary to ensure that adjustment of reboiler duty is not restricted by wrongly setting the fixed variable.

If the reboiler is a fired heater then it may be subject to disturbances in the fuel system. Chapter 10 describes techniques which will keep duty constant.

In addition to the energy balance or material balance options, there are various hybrid schemes – the most well-known of which is the simplified *Ryskamp scheme* (Reference 29). In this version of the scheme we add a reflux ratio controller to the energy balance scheme, as shown in Figure 12.70. The input to the ratio algorithm is the drum level controller output (or the distillate flow measurement). So, as the level controller takes corrective action, it moves reflux and distillate in proportion – maintaining a constant reflux ratio. An advantage of the scheme is that it helps break the interaction between the two composition controllers. Figure 12.71 shows, on an example column, the effect of changing reboiler duty with each of the three level control schemes in place. The composition targets for HK_d and LK_b are both 5%. It confirms that the material balance scheme is better than the energy balance scheme at keeping HK_d constant, but the Ryskamp greatly outperforms both. This means that when the bottoms composition controller adjusts reboiler duty the disturbance made to the distillate composition is almost negligible. The distillate composition controller would, when required to, adjust the reflux ratio. Doing so will still cause a disturbance to the bottoms composition but breaking the interaction in one direction is enough to enable stable composition control. A second advantage of the Ryskamp scheme is that, compared to the material balance scheme, it makes the bottoms composition more sensitive to changes in reboiler duty, as seen in Figure 12.72. This means that less adjustment of reboiler duty is required – reducing the interaction further.

To fully exploit the benefit of the Ryskamp scheme, the partial decoupling it provides needs to be achieved dynamically as well as at steady state. To achieve this the drum level controller needs to be tightly tuned so that the reflux is corrected quickly when the reboiler duty is changed. This of course will also quickly adjust the distillate flow which would be disadvantageous to any downstream unit. This can be resolved by determining, from the drum level measurement, the rate of change of the volume of liquid

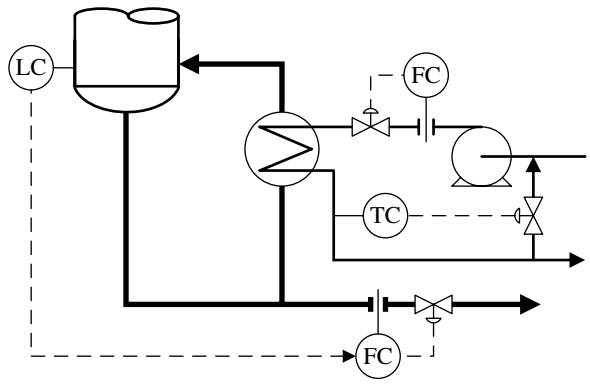

Figure 12.69 *Temperature control of hot oil return*

held in the drum (dV/dt). Suitable calculation methods for this are described in Chapter 4. Using this result and the target R/D, the SP of the reflux flow controller (R_{SP}) is then instead determined from the measured reflux (R) and distillate (D) flows.

$$R_{SP} = \frac{\left(R/D\right)_{target}}{\left(R/D\right)_{target}+1}\left(R+D+\frac{dV}{dt}\right) = \left(\frac{R}{R+D}\right)_{target}\left(R+D+\frac{dV}{dt}\right) \tag{12.59}$$

This would then permit averaging level control to be employed on the drum.

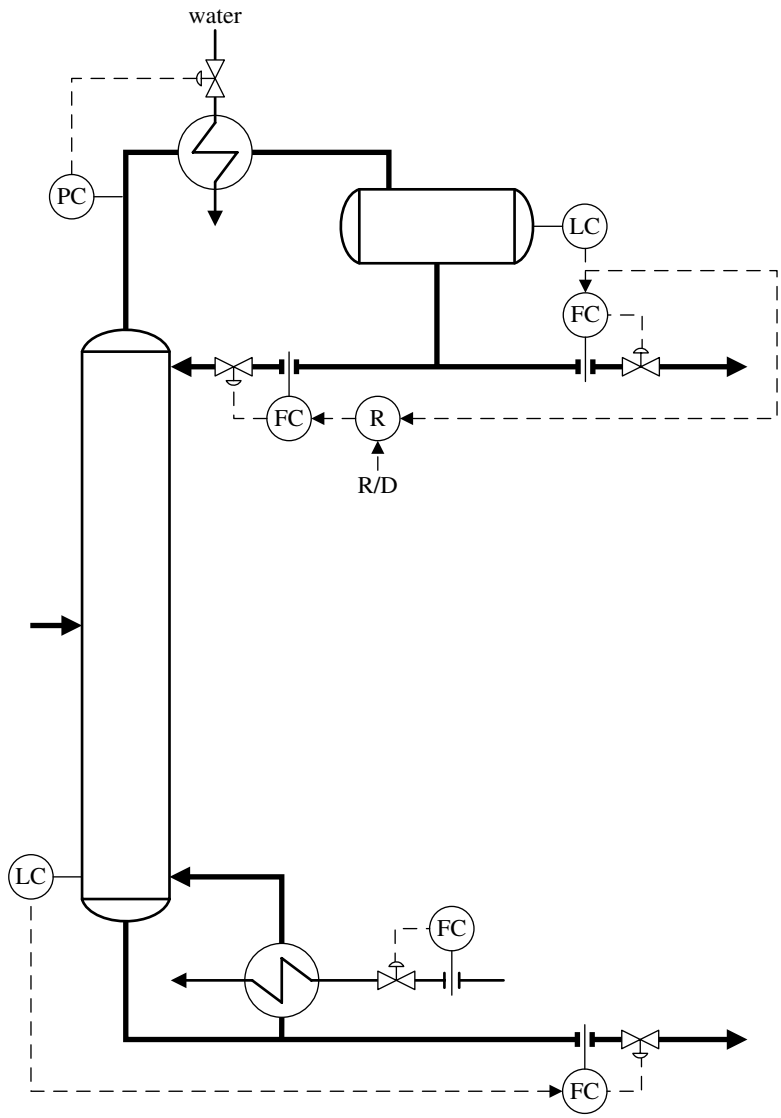

Figure 12.70 Simplified Ryskamp scheme

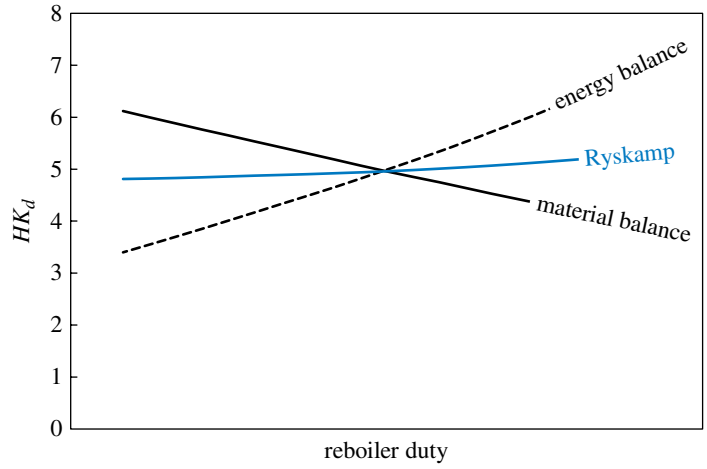

Figure 12.71 *Performance of Ryskamp scheme for distillate composition*

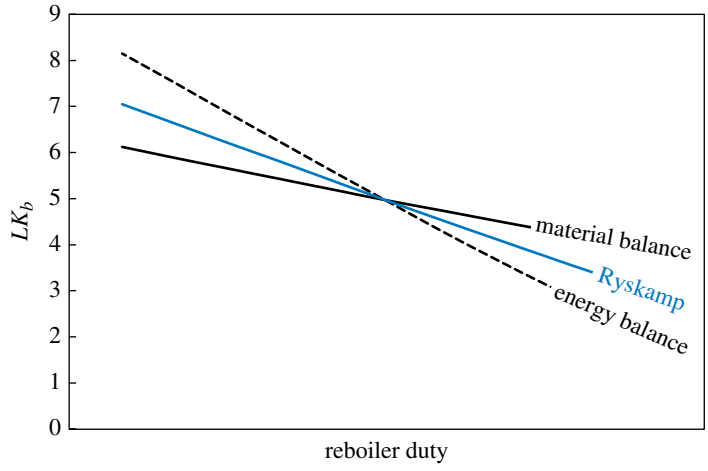

Figure 12.72 *Performance of Ryskamp scheme for bottoms composition*

Of course, because the Ryskamp scheme offers benefits in one operating scenario does not mean it is universally the best scheme. Figure 12.73 shows how each of the schemes perform as feed composition changes. As we expect, the energy balance scheme maintains the distillate composition closer to its target than the material balance scheme. But, despite keeping the reflux ratio constant, the Ryskamp scheme performs poorly. Figure 12.74 explains why. As we know from Equation (12.25) distillate product rate (D) varies linearly with feed composition (LK_f). Perhaps what is not immediately obvious is why the reflux passes through a maximum. To explain this, consider a feed that has a composition identical to the bottoms product specification. No energy is required to operate the column; it effectively becomes the pipework necessary for the feed to all be pumped out as bottoms. No reboil and no reflux are required. Now we consider the other extreme case, where the feed has a composition identical to the distillate product specification. Again no separation is required; we need to provide sufficient reboil duty to fully vaporise the feed,

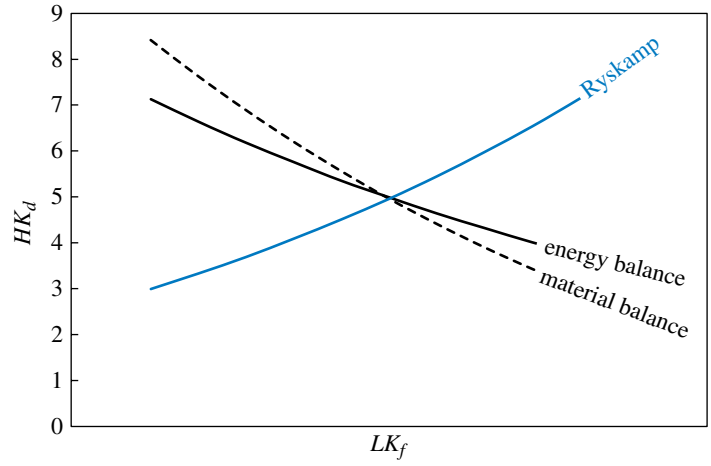

Figure 12.73 *Performance of Ryskamp scheme for changes in feed composition*

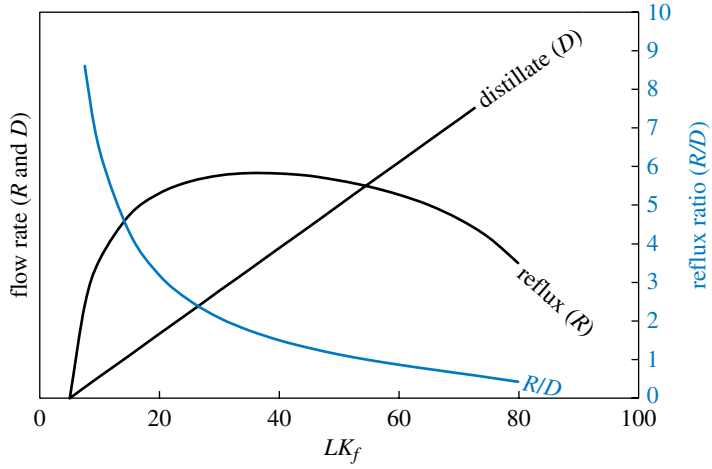

Figure 12.74 *Effect of feed composition*

and sufficient condenser duty to return it to liquid in the reflux drum, but no reflux is required. Figure 12.74 shows that keeping the reflux ratio constant is not what is required. Indeed, the value required varies greatly with feed composition – particularly when the feed contains little of the light key component.

Figure 12.75 shows the original version of the Ryskamp scheme. This maintains constant a different definition of reflux ratio, i.e. $R/(R + D)$. This is chosen because it is the slope of the top operating line on the McCabe-Thiele diagram – see Equation (12.18). The output of the level controller is $R + D$. From this is subtracted R to generate the SP for the distillate flow controller. The reflux ratio target is multiplied by the level controller output to generate the SP for the reflux flow controller. In terms of composition control the performance of this version of the scheme is unlikely to be distinguishable from that shown in Figure 12.70. The additional complexity might therefore not be justified. But, while it is still not correct

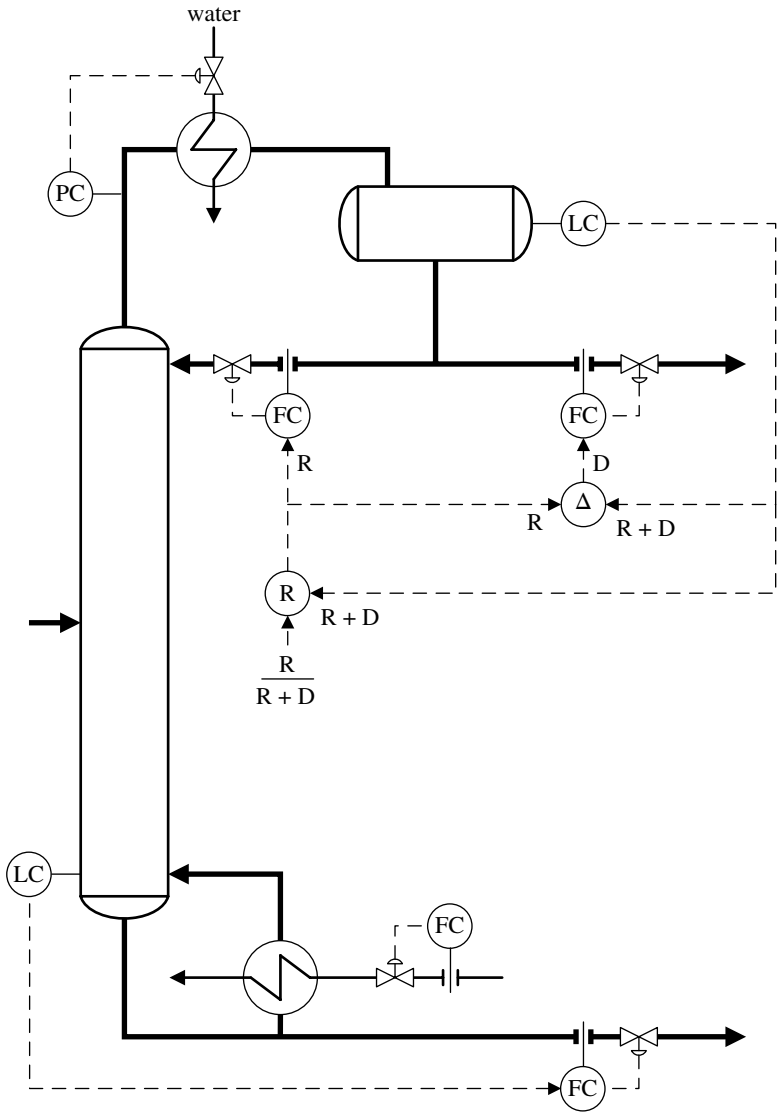

Figure 12.75 *Original Ryskamp scheme*

to keep the modified reflux ratio constant as feed composition changes, Figure 12.76 shows a more linear relationship. This would be helpful if feedforward on feed composition was being contemplated.

Like Ryskamp, a similar scheme can be applied for the column level. Its controller is set up to manipulate the bottoms flow but with its output also used as the input to a reboil-to-bottoms ratio scheme. This will help decouple the bottoms composition from changes made at the top of the column. However there may be issues which adversely affect the level of dynamic decoupling achieved. We have already described the limitations of using reboiler duty to control column level. While using it in combination with bottoms flow will be an improvement, there will remain some restriction on achieving tight level control. Care must also be taken to ensure that reboiler duty is indeed proportional to the manipulated variable. For

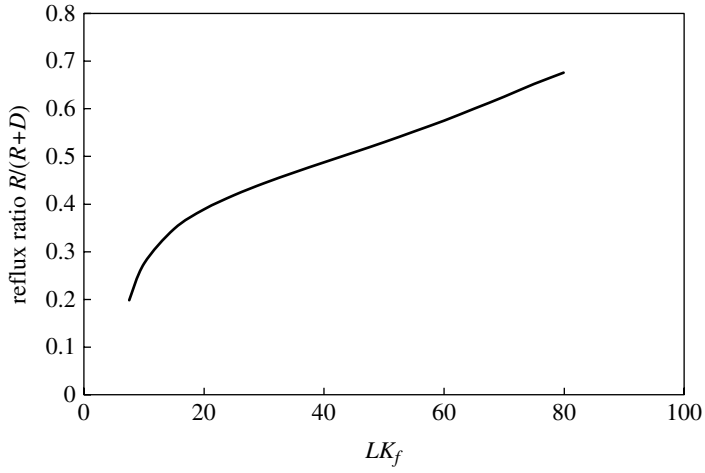

Figure 12.76 *Effect of feed composition on the original Ryskamp scheme*

example, if the heat source is a hot liquid, then reboiler duty will not increase linearly with its flow. We have described, in Figures 12.67 and 12.68, how this might be resolved by installing a duty controller.

Level controller tuning would be determined using the methods described in Chapter 4. As mentioned, tuning must be tight if using the material balance scheme. Averaging tuning can be used for the energy balance scheme if beneficial to minimise flow disturbances to a downstream process. Care must be taken in calculating the volume of the reflux drum. If determined from step-testing, we make the assumption that the drum is in mass balance at the start of the test and the flow into the drum remains constant during the test. This is the case if the distillate flow is stepped. However, if the reflux is sub-cooled, then stepping this flow will change the overhead vapour flow and hence the liquid flow into the drum. The calculated drum volume will then be less than the true volume. However this value should still be used in determining the tuning for the material balance scheme. The true vessel volume should be used in the calculation of tuning for the energy balance scheme.

The tuning of a level controller manipulating reboiler duty cannot be calculated using the methods described in Chapter 4 since these assume we can predict the change in liquid level from the change in manipulated flow. Instead we perform step-testing as described in Chapter 2 and use the resulting dynamics to tune the controller using the methods described in Chapter 3. Because of the lag caused by the reboiler it is common that such level controllers benefit from the inclusion of derivative action. It is also common that they cannot be tightly tuned without becoming unstable.

12.9 Tray Temperature Control

Although on-stream analysers are readily available for many properties they do have limitations. They can be subject to large delays – either due to the analyser technology itself or the position of its sample point. For example the sample point for distillate is commonly located on the discharge of the product pump – usually after any product coolers; thus any change in product composition must first pass through the reflux drum before it is detectable by the analyser. Further, once the disturbance has been detected, we

have a drum full of off-grade material that must pass through the product system. Secondly, although analyser technology is steadily improving, they are more prone to failure than other instrumentation. Some are also required to be taken out of service regularly for calibration and preventative maintenance. And thirdly, they are usually expensive to install and maintain. There may simply not be sufficient benefit to justify the cost.

On many columns tray temperature control offers a method which, although not as accurate as an analyser, provides a degree of composition control and overcomes these problems. It is not necessarily a replacement for a higher level of composition control; if both are feasible then, as we shall show later, they can operate in conjunction. Tray temperature control works on the principle that liquid on the trays is at its bubble point. Bubble point is related to composition and so fixing the bubble point provides some level of composition control. As we have seen cut can also be expressed as temperature and so controlling tray temperature helps maintain cut.

Ideally the control engineer should be involved in the selection of the tray(s) on which any temperature controller will be installed. It is more often the case that existing temperatures, selected by others, have to be used. Retrofitting new temperatures is usually difficult. Most vessels are stress-relieved after construction. This involves heating the vessel to a temperature of around 650°C (1200°F) – depending on the alloy. Doing so avoids stress corrosion cracking when the vessel is in service. The stresses introduced by drilling the vessel and welding to it an additional nozzle and flange (to contain the thermowell) would require this process to be repeated – at least in the area around the weld. With the vessel in place, probably lagged with insulation material, attached to the structure and to pipework, this would be a very costly exercise. We will show later that it is usually beneficial to install additional temperature measurements as inputs to inferential properties. The resulting improvement to the accuracy of the inferential property can be such as to make an analyser unnecessary. It is a false economy to install only the minimum number of temperature measurements.

In theory we should control the temperature at the point at which the product is withdrawn from the column. Indeed, if the column has a side-draw, the measurement is best located just below the draw tray. However, for the distillate and bottoms draws, there are several reasons why this may not be practical. Firstly, the liquid may not be homogeneously at its bubble point. This is likely to be the situation close to the top tray since reflux is often sub-cooled. It can also occur at the base of the column if the vapour from the reboiler is superheated. Secondly, in pseudo-binary columns, the relationship between composition and bubble point also depends on the proportion of non-key components in the product. The greatest proportion of LLK will occur at the top of the column, and of HHK at the bottom. Thus, if non-key composition varies, these regions are most prone to inaccuracy.

When moving away from the top and bottom of the column, there are other issues to consider. If the key components have very similar bubble points it may not be possible to identify a tray anywhere where the temperature is sufficiently sensitive to changes in composition. Figure 12.77 shows the relationship between bubble point and composition for a *C3 splitter* – a column separating propene from propane. A common requirement is that the propene be 99.5% pure. The relationship shows that doubling the permitted amount of propane in the product changes the tray temperature by about 0.4°C (0.7°F) – a change too small to be measured accurately by conventional instrumentation. In such a case it is likely that accurate composition control could only be achieved with the use of an on-stream analyser.

The relationship between composition and temperature may be nonlinear. While this will not affect temperature control it will give tuning problems with any composition controller that may later be cascaded to the temperature controller. Figure 12.78 shows such a problem arising from a poorly selected tray in our LPG splitter. Considering the design case where our target is 5% C_4 in distillate then, provided we are between 4 and 6%, the variation in the slope of the relationship corresponds to an acceptable variation in process gain of about ±20%. However, if a disturbance were to take the composition down to around

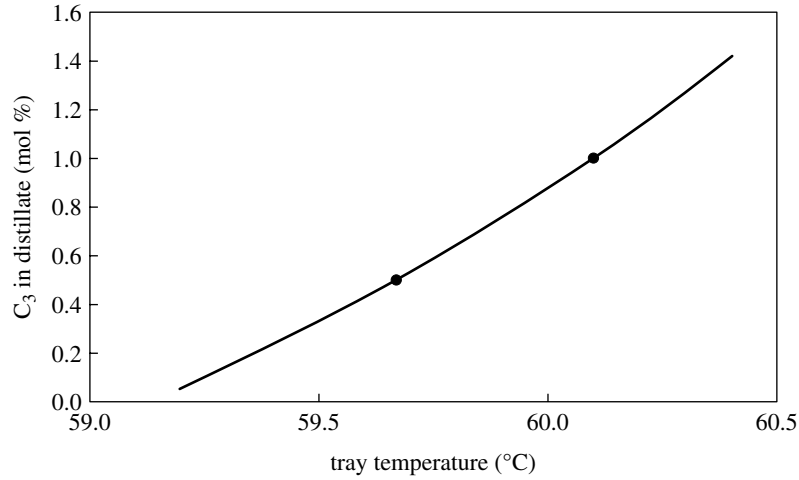

Figure 12.77 *Insensitivity of temperature to composition in C3 splitter*

1%, the process gain falls by a factor of around 10 – meaning that the composition controller would be very slow in returning to target. Similarly the relationship between temperature and its MV may be non-linear which will create tuning problems for the temperature controller. We will address this later when completing the exercise of checking the suitability of a chosen tray temperature.

There is also potentially a problem in locating the temperature too close to the feed tray. On the tray itself it is unlikely that liquid will be homogeneously at its bubble point. There is also an issue with trays close to the feed tray because the relationship between tray temperature and composition can be sensitive to feed composition. The dashed curve in Figure 12.78 shows another reason why the tray is a poor choice. A 5% reduction in the LK content of the feed causes a major change in the relationship between distillate composition and tray temperature. If we were controlling the tray at around 70°C to meet our 5% target and the feed composition were to change, then our product composition would move to around 1.4%. Figure 12.79 shows temperature profiles drawn for two different feed compositions. In both cases the products are exactly at target composition. The profiles show that, if we were to fix the temperature at any point in the column, at least one of the product compositions would change if feed composition changed. As one might expect, the nearer to the feed tray that we control the temperature the more pronounced the problem.

The process of selecting tray temperature(s) for composition control is usually completed using a simulation of the column, rather the real column. Properly it should be done before the column design is frozen. On columns already built, that have multiple tray temperatures installed, it is possible to execute the plant tests required but they will be very disruptive to the operation.

The column at this stage of the control design will have two MVs remaining for use to control the composition of both products. What these variables are depends on the choice of level control configuration. If the material balance scheme is in place then, usually, distillate flow and reboil are available or, less usually, bottoms flow and reflux. If the energy balance scheme is in place then reboil and reflux are available.

On a 20 tray column, with the energy balance scheme in place, Figure 12.80 shows the effect of varying reboil duty by ±5%. Note that profile translates horizontally; all tray temperatures increase with reboil duty. If the material balance scheme were in place, increasing reboil duty would increase separation

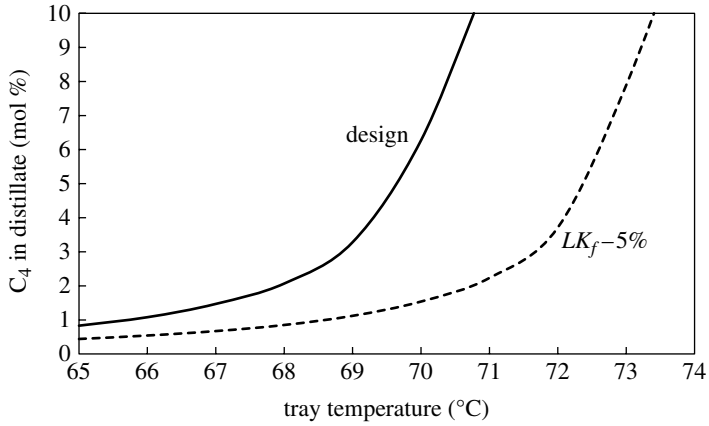

Figure 12.78 *Non-linear relationship between composition and temperature*

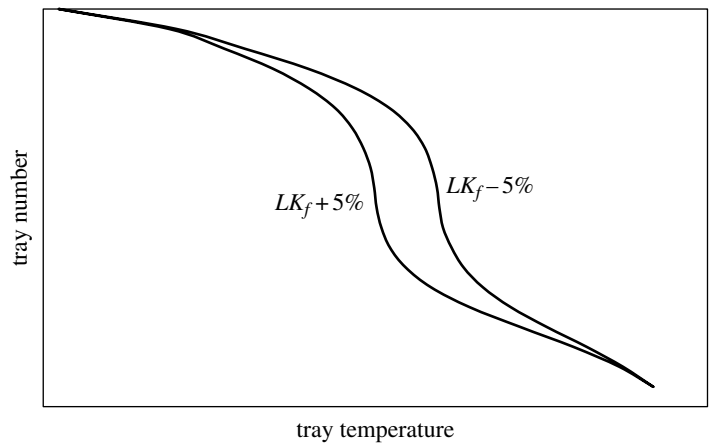

Figure 12.79 *Effect of feed composition on temperature profile*

without changing product yields. Thus both products approach the bubble point of the pure LK and HK components and the temperature profile rotates (anti-clockwise). This behaviour forms the basis of a potential decoupling technique that we cover later in this chapter.

The profiles in Figure 12.80 have been plotted in Figure 12.81 as changes in temperature. We would likely wish to manipulate reboiler duty to control the bottoms composition and so we would select a tray in the lower section of the column. In choosing a suitable tray we look for sensitivity and linearity. Sensitivity is measured by the distance between the two profiles and linearity by the symmetry. We can see that tray 4 shows the greatest sensitivity. In terms of linearity an increase in reboil duty has an effect about 50% greater than a decrease. While nonlinear, it falls just within our criterion of being able to select a value for process gain which varies by less than 20%. Had it not, then we might reconsider our choice. For example tray 2, although considerably less sensitive, shows almost exactly linear behaviour. Tray 3 might be considered as a good compromise.

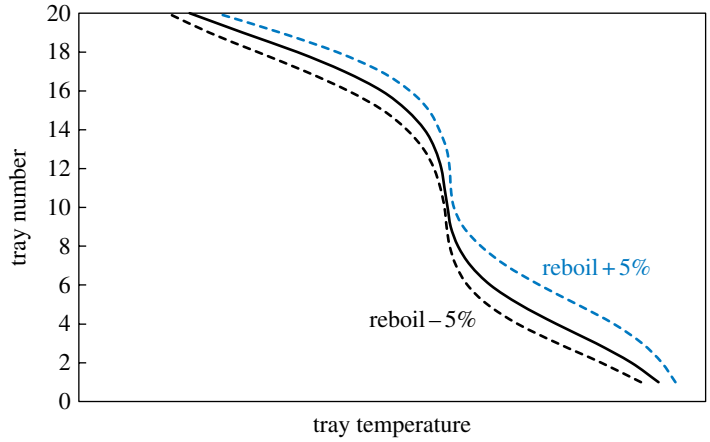

Figure 12.80 *Effect of changing reboil duty (with energy balance scheme in place)*

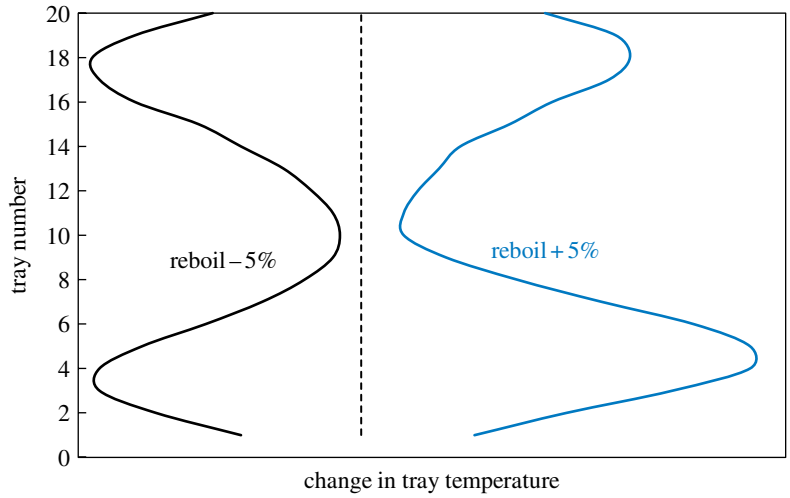

Figure 12.81 *Change in temperature due to reboil*

Figure 12.82 shows, in a different form, the sensitivity and linearity of trays 3 and 4. The curves have also been plotted also to show the effect of changing feed composition. If we were to install a temperature controller on tray 4, initially holding the composition at 5%, then the change in feed would cause the composition to increase by about 0.8%. Using tray 3, since it is further from the feed tray, the disturbance is reduced to about 0.5%. Taking the better linearity into account, then tray 3 is probably the better choice.

While we have confirmed there is sufficient linearity between composition and temperature our first concern is that we can control temperature by manipulating reboil duty. This check is shown in Figure 12.83 which shows how the temperatures on tray 3 and 4 vary. While both would work well, tray 3 shows greater linearity while tray 4 shows greater sensitivity.

Having configured the lower temperature controller we can complete a similar exercise for the upper controller. Figure 12.84 shows how the profile changes if we change reflux by ±5%; Figure 12.85 shows

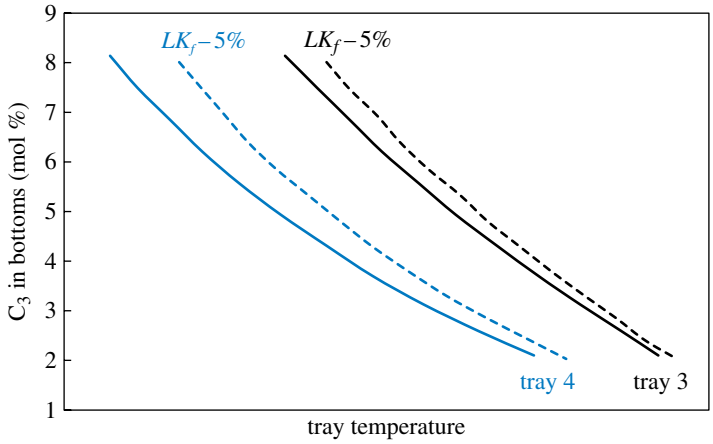

Figure 12.82 *Effect of feed composition on lower tray temperature*

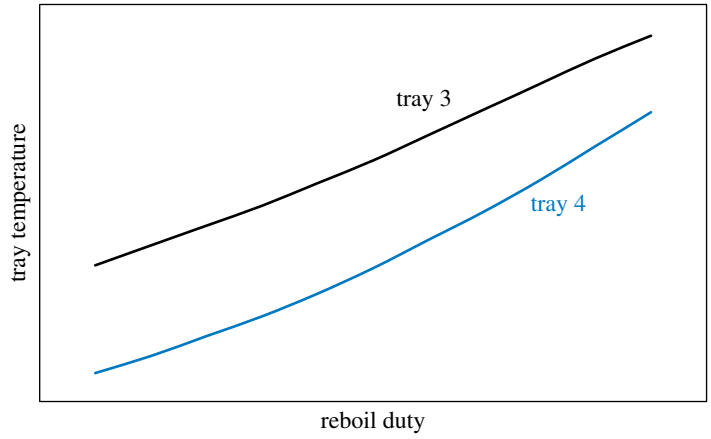

Figure 12.83 *Linearity of lower temperature control*

the changes in temperature. Reflux would typically be manipulated to control overhead composition and we choose a tray in the upper section of the column. Here we can see that tray 18 shows the greatest sensitivity. An increase in reflux causes a change about 40% greater than a decrease – within our criterion for linearity. If it were not then tray 19 shows greater linearity with slight loss of sensitivity but might be rejected because it is too close to the reflux tray above. Tray 17, although of similar sensitivity to tray 18, is well outside our linearity criterion with an increase in reflux causing a temperature change around 75% larger than a decrease.

Figure 12.86 shows the effect that a change in feed composition would have on composition control. Tray 18 remains the preferred choice, since it is further from the feed tray. Fixing the temperature to initially hold the composition at 5% would result in it changing by about 0.4% when the feed composition changes. Figure 12.87 shows that tray 18 is readily controllable by manipulating reflux.

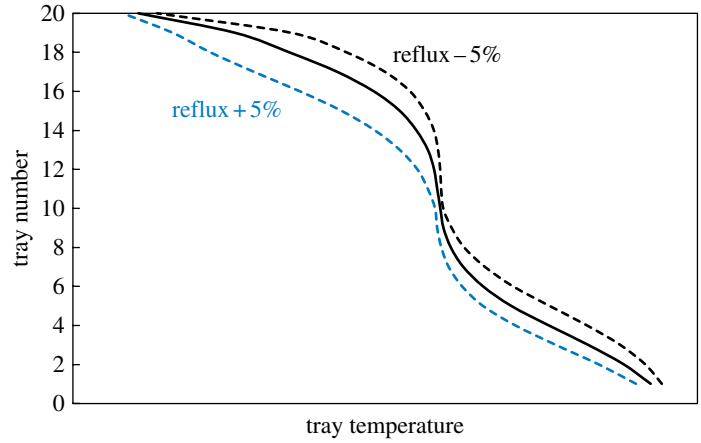

Figure 12.84 *Effect of changing reflux (with energy balance scheme in place)*

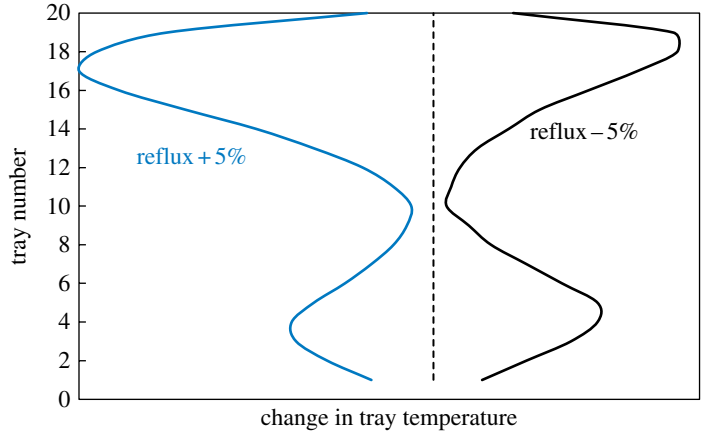

Figure 12.85 *Change in temperature due to reflux*

It can be advantageous to use multiple tray temperature measurements in the same controller. There may not a measurement on the required tray but it may be possible to infer a value by interpolating between two or more other measurements. Or a temperature on one tray may become insensitive to composition changes under different operating conditions; incorporating a second measurement that correspondingly becomes more sensitive will maintain composition control.

If T_1 is the temperature measured on tray n_1, T_2 is the temperature measured on tray n_2 and T is the temperature on the required tray n, then the equation of the section of the temperature profile connecting them, assuming a straight line, is

$$\frac{T - T_1}{T_2 - T_1} = \frac{n - n_1}{n_2 - n_1} \tag{12.60}$$

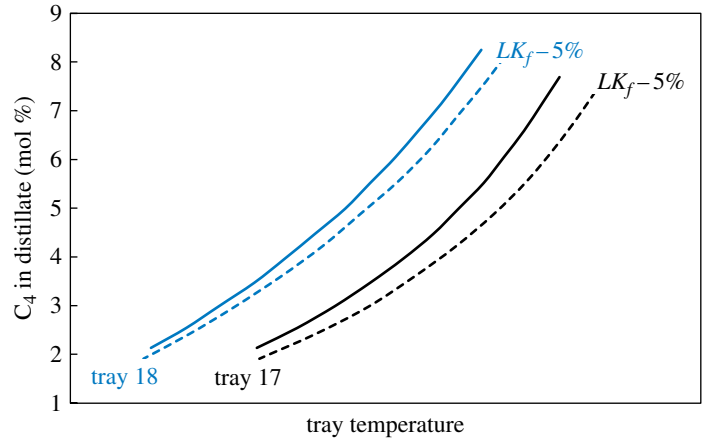

Figure 12.86 *Effect of feed composition on upper tray temperature*

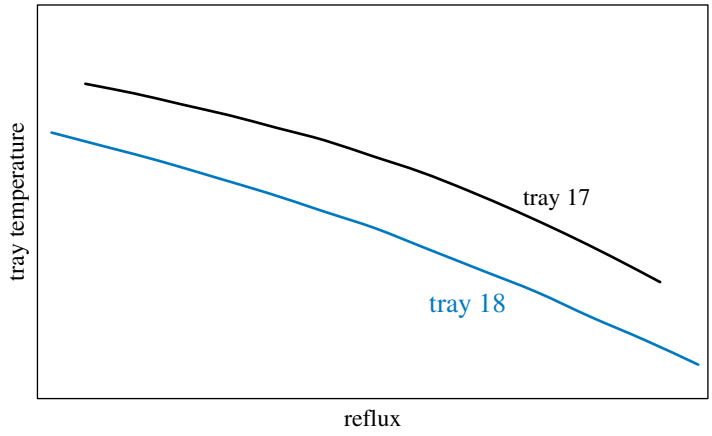

Figure 12.87 *Linearity of upper temperature control*

Rearranging

$$T = \frac{n_2 - n}{n_2 - n_1} T_1 + \frac{n - n_1}{n_2 - n_1} T_2 \tag{12.61}$$

This is a weighted average of the two measured temperatures with weighting factors designed to give the temperature on a tray where the required measurement does not exist. If multiple measurements are used to maintain sensitivity, then the weighting should be chosen to keep constant the process gain between composition and temperature. Unusual dynamic behaviour may arise if the tray temperatures are far apart. Any resulting inverse response may require the controller to be so de-tuned as to counter the advantage of using multiple measurements.

Equation (12.60) can be rearranged

$$n = \frac{T_2 - T}{T_2 - T_1} n_1 + \frac{T - T_1}{T_2 - T_1} n_2 \tag{12.62}$$

This is useful if the temperature profile is very steep, for example if temperature is insensitive to composition throughout the column. T then becomes the required temperature and n the tray on which this temperature target is currently being met. The value of n is used as the PV of a controller with the required tray as its SP. This can be confusing to the operator; to resolve this we also apply Equation (12.62) to convert the operator-entered temperature SP to the SP for n.

At this stage we should emphasise that we are not yet in a position to commission both lower and upper temperature controllers. From our understanding of the process we know that the two controllers will interact – possibly to the point of becoming unstable. We will address this later. Until then we will proceed on the basis that only one of the controllers will be in service.

A very common configuration is to place a single temperature controller (TC) in the lower section of the column – manipulating reboiler duty. Figure 12.88 shows the effect of doing so with the energy balance scheme in place. During feed rate changes, with no TC, both product compositions will clearly vary. Figure 12.88 shows the impact on the distillate composition if reflux (R) and reboil (V) are held constant (coloured line). It also shows that, if the TC is commissioned to manipulate the reboiler duty, while this will maintain the bottoms composition almost constant it does so at the cost of worsening the variability of the distillate composition (dashed line). To compensate for this reflux must be adjusted. Without a second TC in place one possible solution is to maintain reflux in proportion to feed rate (solid black line). Figure 12.89 shows that making feed rate changes with the material balance scheme in place gives the same problem and that this can be resolved by keeping distillate flow in proportion to feed. The conclusion is therefore that implementing temperature control in the lower section of the column causes the distillate composition to vary considerably more when column feed rate changes – no matter which level control strategy is adopted. This can be resolved by the implementation of ratio feedforward to the remaining unused MV.

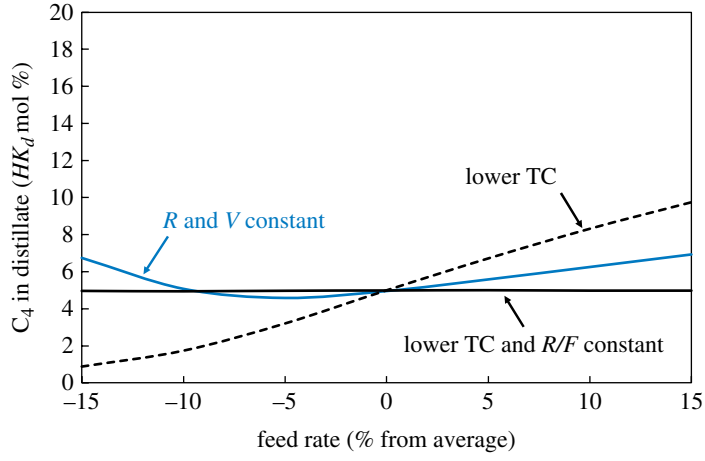

Figure 12.88 *Impact of lower temperature control (energy balance scheme)*

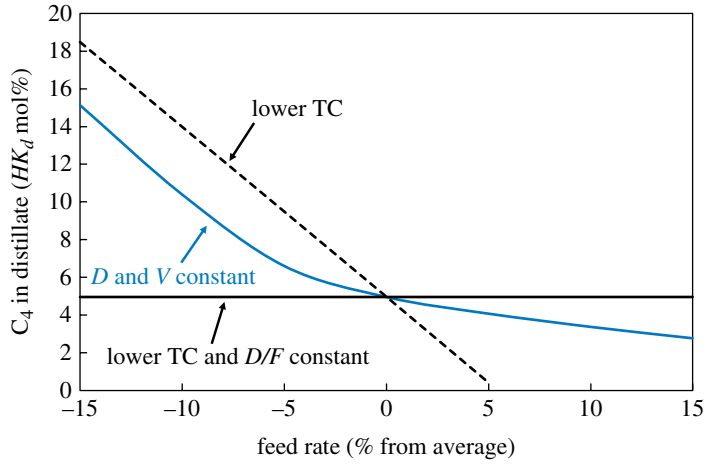

Figure 12.89 *Impact of lower temperature control (material balance scheme)*

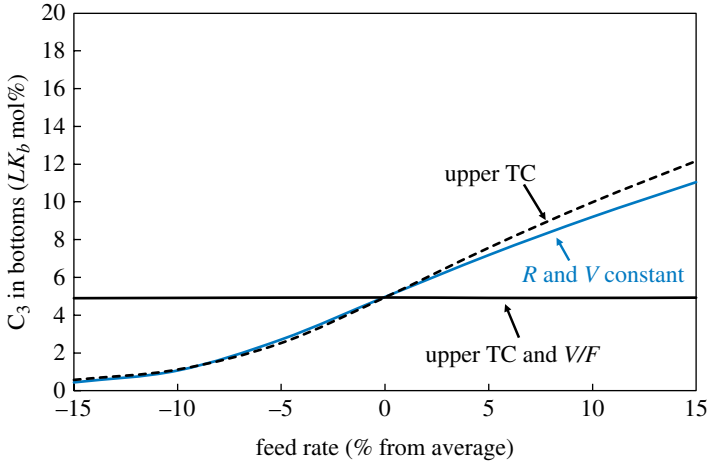

Figure 12.90 *Impact of upper temperature control (energy balance scheme)*

Figures 12.90 and 12.91 show the same tests but this time with a TC commissioned in the upper section of the column. In the case of the energy balance scheme it manipulates reflux; if the material balance scheme is in place it manipulates distillate flow. In both cases controlling the distillate composition has much less of an effect on the bottoms composition. This demonstrates the importance of cut versus separation as a means of controlling composition. The TC in the upper section manipulates distillate flow, either directly or indirectly via reflux, and keeps *D/F* approximately constant. Not compensating fractionation has less of an impact. However, while the upper TC does not significantly worsen the control of bottoms composition, maintaining a constant *V/F* ratio is still very beneficial.

Manipulating cut should therefore the first choice if our aim is to control only one of the product compositions. While it is quite feasible to adopt this approach if the TC is close to the top of the column, a TC in the lower section is likely to respond slowly to changes in reflux and therefore may not control well

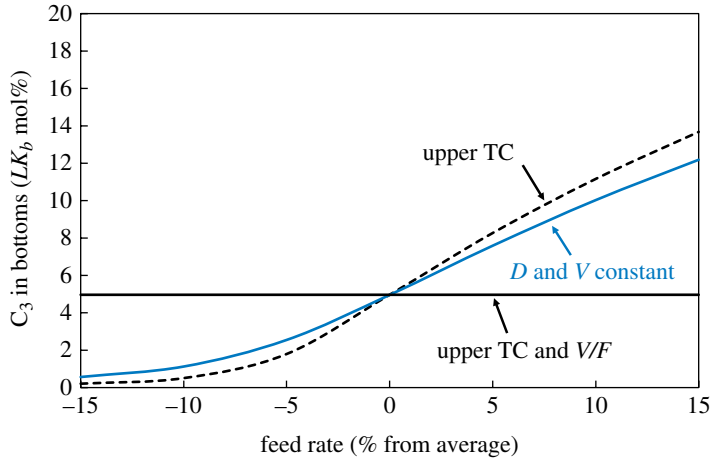

Figure 12.91 *Impact of upper temperature control (material balance scheme)*

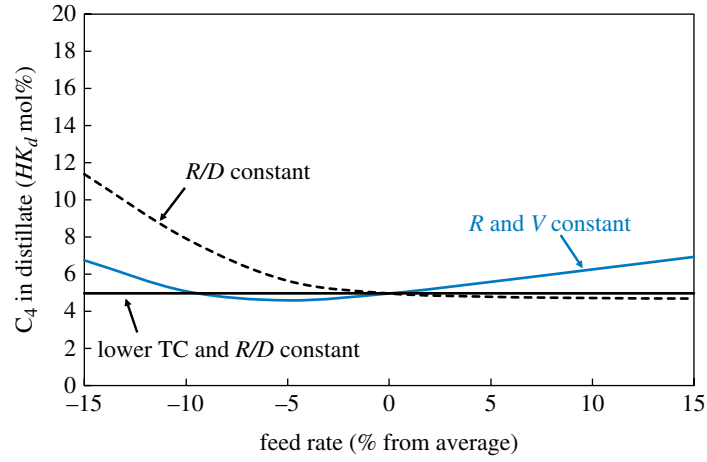

Figure 12.92 *Impact of lower temperature control (Ryskamp scheme)*

during process disturbances. If the less preferred version of the material balance scheme is in place, a TC in the lower section might be thought to respond quickly since it would manipulate bottoms flow. However, this will not be the case because of the lag introduced by the reboiler responding to the change in column level.

Maintaining *R/F*, *V/F*, *D/F* (or possibly *B/F*) constant are feedforward strategies that we will address later in this chapter. While the Ryskamp scheme would appear to include some feedforward (in that it keeps *R/D* constant) it should be used with care. Figure 12.92 shows that the Ryskamp scheme commissioned without a TC in the lower section of the column gives poorer control of distillate composition than the energy balance scheme. With the TC in place it performs extremely well. In designing the scheme thought should be given to automatically disabling the *R/D* ratio controller if the TC is switched to manual (and re-commissioning it when the TC is returned to auto).

12.10 Pressure Compensated Temperature

By selecting the optimum trays we have dealt with many of the problems that can arise with tray temperature control. There remain some other issues that we need to address with the controller design. The first of these is the effect of column pressure. The relationship between composition and bubble point depends on pressure, as shown in Figure 12.93. To maintain a composition of 5% C_4 in distillate, when operating at normal (*reference*) pressure, the tray temperature would be controlled at T. If the column pressure is reduced by ΔP, then the bubble point will reduce and the correlation between composition and bubble point changes. Maintaining the tray temperature constant will result in the C_4 in distillate increasing to around 6.7%. To maintain a constant composition the tray temperature should be reduced by ΔT. Figure 12.94 shows similar behaviour in the lower section of the column, where the reduction in pressure

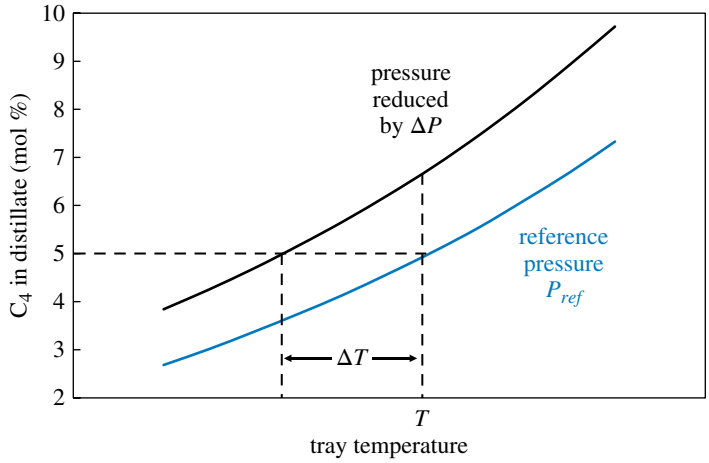

Figure 12.93 *Effect of pressure on distillate composition*

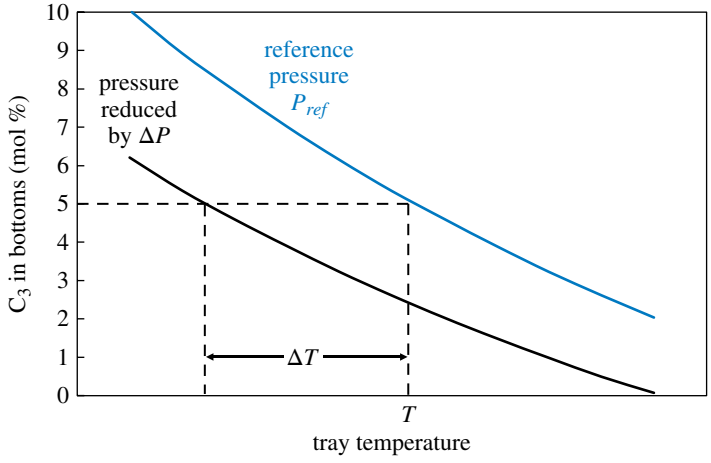

Figure 12.94 *Effect of pressure on bottoms composition*

causes the C_3 in bottoms to fall to about 2.4% – requiring a reduction in tray temperature greater than that required in the upper section.

The technique for dealing with this problem is known as *pressure compensated temperature (PCT)*. This was covered briefly as an example of signal conditioning in Chapter 5. In its simplest form a linear correction is applied to the measured tray temperature (T) based on the measured pressure (P).

$$PCT = T - \frac{dT}{dP}\left(P - P_{ref}\right) \tag{12.63}$$

Taking our example of a reduction in pressure, the value of *PCT* will be higher than *T*. If we make *PCT* the PV of the tray temperature controller, then it will compensate for the increase by reducing *T*. Provided we have quantified *dT/dP* correctly for each temperature controller (as $\Delta T/\Delta P$), then the product compositions will remain constant.

It is possible, although unusual, to apply the pressure correction to the SP of the tray temperature controller (T_{SP}). Some operators prefer this since it emulates what they would do when the pressure changes. Further the controller still displays the real tray temperature.

$$SP = T_{SP} + \frac{dT}{dP}\left(P - P_{ref}\right) \tag{12.64}$$

We need to be cautious about the value that we use for *P*. Consider the situation where the column is operating at the condenser duty limit and the pressure controller is close to saturation. If we use the PV of the controller in the calculation of *PCT*, then a transient increase in pressure will cause the lower tray temperature controller to increase the reboiler duty. This will overload the condenser more, causing a further increase in pressure and a further increase in reboiler duty. In these circumstances it would be better to use the SP of the column pressure controller in the calculation of *PCT*. This also has the advantage of being noise free. If the pressure controller is switched to manual, and PV tracking is implemented, the PCT calculation will automatically use the PV. If this is permitted, then it may be necessary to filter the measurement and add some logic to prevent the compensation driving the unit towards the condenser limit. This situation can also arise if a column has multiple pressure measurements at different locations – one of which will be used for control but we may wish to use another for calculating *PCT*.

Theoretically the pressure transmitter should be at the same location in the column as the temperature transmitter. Fortunately on most columns the pressure difference, between the tray and the point at which column pressure is measured, is small compared to the operating pressure. Secondly, we base *PCT* on the change in pressure which, provided the column pressure drop remains constant, will be the same throughout the column. We need to be cautious, however, on vacuum columns where changes in pressure drop can be of a similar order of magnitude to the column pressure. We will see later that the value of *dT/dP* in this type of column is extremely large and will so amplify any error in determining the change in pressure.

There are several ways in which *dT/dP* might be determined. The first assumes that sufficient good quality historical data exist so that an inferential property can be developed in the form

$$Q = a_0 + a_1 T + a_2 P \tag{12.65}$$

Differentiating

$$dQ = a_1 dT + a_2 dP \tag{12.66}$$

As P changes we want Q to remain constant or dQ to be zero, and so

$$\frac{dT}{dP} = -\frac{a_2}{a_1} \qquad \text{and} \qquad PCT = T + \frac{a_2}{a_1}\left(P - P_{ref}\right) \tag{12.67}$$

If historical data are not available, then another approach is plant testing. This entails operating at different pressures and either allowing any composition controllers to take corrective action or manually adjusting conditions to bring the compositions back to target. Collecting data at several different pressures will allow tray temperature(s) to be plotted against pressure. The slope of the line is then *dT/dP*.

There are several theoretical approaches to determining *dT/dP*. We have already used the Antoine Equation (12.1) to estimate relative volatility. Differentiating we get

$$\frac{dP}{P} = \frac{B.dT}{\left(T + C\right)^2} \tag{12.68}$$

By substituting for T, again using the Antoine Equation, *dT/dP* can be defined as a function of either T and P or only P.

$$\frac{dT}{dP} = \frac{\left(T + C\right)^2}{B.P} = \frac{B}{P\left(A - \ln\left(P\right)\right)^2} \tag{12.69}$$

Note that, if the version of the Antoine Equation being used is based on $\log_{10}(P)$ rather than $\ln(P)$, then this can be replaced before differentiation.

$$\log_{10}\left(P\right) = \log_{10}\left(e\right) \times \ln\left(P\right) = \frac{\ln\left(P\right)}{\ln\left(10\right)} = 0.4343 \times \ln\left(P\right) \tag{12.70}$$

Assuming a typical operating pressure of 12 barg (an absolute pressure of 13.01325 bara or 12.84308 atm) for a LPG splitter, and the Antoine coefficients given in Table 12.2, we obtain the values for *dT/dP* shown in Table 12.7.

An alternative to the Antoine Equation is the Clausius-Clapeyron Equation which predicts *dT/dP* directly from the heat of vaporisation (λ).

$$\frac{dT}{dP} = \frac{RT^2}{\lambda.MW.P} \tag{12.71}$$

Table 12.7 *Pressure compensation factors from Antoine Equation*

Component		*dT/dP*	
		°C/bar	°F/atm
propene	C_3H_6	3.27	5.97
propane	C_3H_8	3.39	6.18
n-butane	C_4H_{10}	3.93	7.17

R is the Universal Gas Constant and has a value of 8.314 kJ/kg-mole/K (1.9859 BTU/lb-mole/°F). Unlike the Antoine Equation the temperature T is on an absolute basis. λ should have units of kJ/kg (or BTU/lb). The pressure (P), like Antoine, is on an absolute basis.

To apply Clausius-Clapeyron, we would normally use the actual tray temperature as the value for T. In the absence of this, its value is estimated by applying the Antoine Equation and converting to an absolute basis. Indeed, this comparison between predicted and measured tray temperature can normally be used to confirm that the coefficients used in the Antoine Equation are correct. Table 12.8 shows that the values of dT/dP derived are very close to those generated from the Antoine Equation.

However, the Antoine and Clausius-Clapeyron Equations are restricted to pure components. Antoine coefficients are published for pure components and the Clausius-Clapeyron Equation includes the molar heat of vaporisation – also only known for pure components. While we could assume that the estimates for dT/dP blend linearly with composition, we would need to know the composition on the tray on which the temperature is controlled. A simple approximation is to assume that dT/dP varies linearly up the column between the value for the HK component and that for the LK. So, for example, based on the Antoine results the value for tray 3 is given by

$$\frac{dT}{dP} = 3.93 - \frac{3}{20}\left(3.93 - 3.39\right) = 3.85 \tag{12.72}$$

And that for tray 18 is given by

$$\frac{dT}{dP} = 3.93 - \frac{18}{20}\left(3.93 - 3.39\right) = 3.44 \tag{12.73}$$

While clearly a very simplistic approach, dT/dP varies relatively little between the pure key components and so the error introduced by the approximation is small. However, the underlying assumption that dT/dP is constant is likely to introduce a much larger error. Equations (12.69) and (12.71) both show that dT/dP varies with pressure. Figure 12.95 shows, for values derived from the Antoine Equation, what error is potentially introduced.

On most columns it would not be possible for the pressure to be varied over a wide range before reaching equipment limits. For example, on the case study column, a 1 bar variation in column pressure would change the true value of dT/dP by about 0.2°C/bar (around 0.4°F/atm) – introducing an error of about 5% into the temperature correction. In these circumstances a more rigorous approach is probably not justified. However, there are columns which have multiple operating modes between which pressure may change greatly.

Table 12.8 *Pressure compensation factors from Clausius-Clapeyron Equation*

Component		T (from Antoine)		MW	H_v		dT/dP	
		K	°R		kJ/kg	BTU/lb	°C/bar	°F/atm
propene	C_3H_6	303.5	546.3	42.08	439.43	188.92	3.18	5.81
propane	C_3H_8	311.8	561.2	44.10	425.68	183.01	3.31	6.03
n-butane	C_4H_{10}	366.3	659.4	58.12	385.95	165.93	3.82	6.97

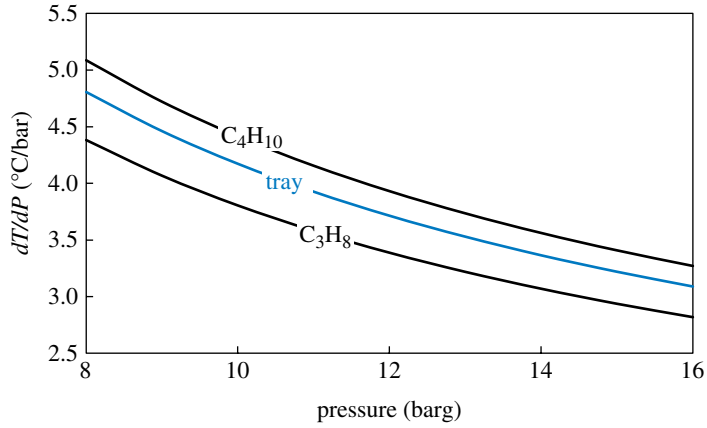

Figure 12.95 *Variation of pressure compensation factor with pressure (Antoine)*

The Antoine and Clausius-Clapeyron Equations can be adapted to determine *PCT* directly. Rearranging the Antoine Equation (12.1)

$$T = \frac{B}{A - \ln(P)} - C \tag{12.74}$$

Writing it for reference conditions

$$PCT = \frac{B}{A - \ln(P_{ref})} - C \tag{12.75}$$

Combining Equations (12.74) and (12.75)

$$PCT = T - B\left[\frac{1}{A - \ln(P)} - \frac{1}{A - \ln(P_{ref})}\right] \tag{12.76}$$

Rearranging the Clausius-Clapeyron Equation (12.71)

$$\frac{dT}{T^2} = \frac{R}{\lambda.MW} \times \frac{dP}{P} \tag{12.77}$$

Integrating

$$\left[\frac{-1}{T}\right]_{T}^{PCT} = \left[\frac{R}{\lambda.MW}\ln(P)\right]_{P}^{P_{ref}} \tag{12.78}$$

Remembering that both T and PCT are absolute temperatures

$$PCT = \frac{1}{\frac{R}{\lambda.MW}\ln\left[\frac{P}{P_{ref}}\right] + \frac{1}{T}} \tag{12.79}$$

While addressing the issue of the variability of dT/dP with pressure, Equations (12.76) and (12.79) still assume pure components. To apply them we have to make two assumptions. The first is to make some estimate of the composition of the liquid on the tray. The second is that, for the Antoine approach, the coefficients A and B can be derived by taking a weighted average of the values for the LK and HK components. With the Clausius-Clapeyron approach the assumption that λ and MW can be derived in this way is perhaps more correct.

A method which needs make no assumption about composition is based on charts developed by Maxwell and Bonnell [30]. The method is published in imperial units and so pressure (P) must be converted to an absolute basis in units of mm Hg and the temperature (T) to an absolute basis in °R. The method will at first appear complex. However the end result is effective and simple to implement. The calculation can readily be implemented in a spreadsheet. This helps avoid the almost inevitable errors that will arise from using a calculator and provides a method that can quickly be applied to any column.

Values for the coefficients A, B, C and D are selected from Table 12.9. The values are used to calculate χ.

$$\chi = \frac{A - B\log_{10}(P)}{C - D\log_{10}(P)} \tag{12.80}$$

Differentiating

$$\frac{d\chi}{dP} = \frac{D\left(A - B\log_{10}(P)\right) - B\left(C - D\log_{10}(P)\right)}{P\left(C - D\log_{10}(P)\right)^2 \ln(10)} \tag{12.81}$$

The normal boiling point of the liquid is calculated from

$$T_b' = \frac{748.1\chi}{\frac{1}{T} - 0.0002867 + 0.2145\chi} \tag{12.82}$$

Differentiating

$$\frac{d\chi}{dT} = \frac{-T_b'}{\left(748.1 - 0.2145T_b'\right)T^2} \tag{12.83}$$

Combining Equations (12.81) and (12.83)

$$\frac{dT}{dP} = \frac{\left(\frac{d\chi}{dP}\right)}{\left(\frac{d\chi}{dT}\right)} \tag{12.84}$$

Table 12.9 *Coefficients for Maxwell and Bonnell calculations*

Pressure	A	B	C	D
$P < 2$	6.761560	0.987672	3000.538	43.00
$2 < P < 760$	5.994296	0.972546	2663.129	95.76
$P > 760$	6.412631	0.989679	2770.085	36.00

The result for *dT/dP* will have units °R/mm Hg and should be converted as required. While the method is arithmetically complex it only requires the typical operating pressure (P_{ref}) and the normal tray temperature. It requires no knowledge of composition.

The method can be extended to accommodate the variation in *dT/dP* with pressure. Figure 12.96 shows this variation. To incorporate this we calculate *dT/dP*, at the normal tray temperature, at a number of different pressures within the expected operating range (shown as individual points in Figure 12.96). The values for *dT/dP* are plotted against the reciprocal of the absolute pressure, as shown in Figure 12.97. The resulting points will lie close to a straight line. Indeed, this is consistent with the differentiated form of the Antoine Equation (12.69) and the Clausius-Clapeyron Equation (12.71) which both show that *dT/dP* is inversely proportional to absolute pressure. The slope (m) and its intercept (c) on the vertical axis can be used to predict *dT/dP* from the operating pressure (P) in absolute units.

$$\frac{dT}{dP} = \frac{m}{P} + c \tag{12.85}$$

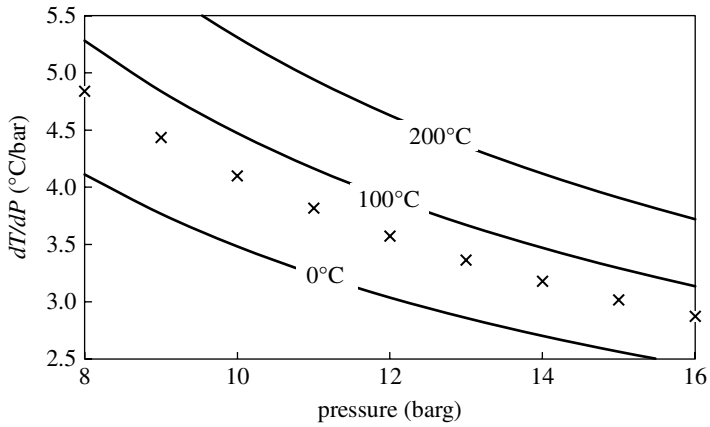

Figure 12.96 *Variation of pressure compensation factor with pressure (Maxwell)*

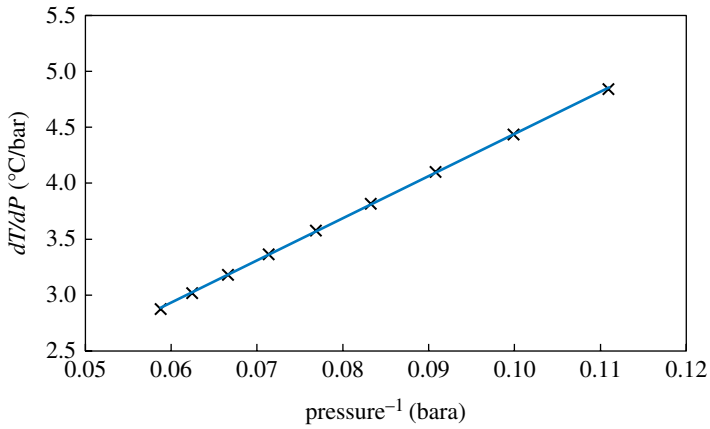

Figure 12.97 *Adjusting pressure compensation factor for pressure*

Integrating

$$\left[T\right]_{T}^{PCT} = \left[m.\ln\left(P\right) + cP\right]_{P}^{P_{ref}} \tag{12.86}$$

Thus

$$PCT = T - c\left(P - P_{ref}\right) - m.\ln\left[\frac{P}{P_{ref}}\right] \tag{12.87}$$

In the same way that *dT/dP* can be determined from a regressed linear inferential of the form of Equation (12.65), so it can be for the nonlinear version. By adding ln(*P*), remembering that *P* is an absolute pressure, we obtain by regression

$$Q = a_0 + a_1 T + a_2 P + a_3 \ln\left(P\right) \tag{12.88}$$

Differentiating

$$dQ = a_1 dT + a_2 dP + \frac{a_3}{P} dP \tag{12.89}$$

As *P* changes we want *dQ* to be zero and so

$$\frac{dT}{dP} = -\frac{a_3}{a_1}\frac{1}{P} - \frac{a_2}{a_1} \tag{12.90}$$

Comparison with Equation (12.85) and substituting in Equation (12.87) gives

$$PCT = T + \frac{a_2}{a_1}\left(P - P_{ref}\right) + \frac{a_3}{a_1}\ln\left[\frac{P}{P_{ref}}\right] \tag{12.91}$$

Those checking the reference from where this technique was developed will find mention that it should only be applied to pure hydrocarbons and narrow-boiling range petroleum fractions. However, this restriction applies primarily to the conversion between the true normal boiling point (T_b) and the boiling point corrected to a *Watson K* of 12 (T_b'). The procedure for calculating *PCT* does not involve this conversion and experience shows that the resulting formula works well.

While not of importance here, for background information, the Watson K factor is used in the oil industry as means of characterising the paraffinicity of a mixture of hydrocarbons and is derived from the *molar average boiling point (MABP)* (in °R) and the specific gravity at 60°F (*SG*).

$$K = \frac{\sqrt[3]{MABP}}{SG} \tag{12.92}$$

If the true normal boiling point is required, for example to validate the method by comparison with a known boiling point, then the procedure is as follows. If *P* is less than 760 mm Hg or the normal boiling point is above 400°F then

$$T_b = T_b' + 2.5\left(K - 12\right)\log_{10}\left(\frac{P}{760}\right) \tag{12.93}$$

If P is greater than 760 mm Hg and the normal boiling point is below 200°F, then

$$T_b = T_b' \tag{12.94}$$

If P is greater than 760 mm Hg and the normal boiling point is between 200°F and 400°F, then

$$T_b = \frac{T_b' - 8.24625(K-12)\log_{10}\left(\frac{P}{760}\right)}{1 - 0.0125(K-12)\log_{10}\left(\frac{P}{760}\right)} \tag{12.95}$$

The one remaining limitation of the method based on Maxwell-Bonnell, as in those based on Antoine and Clausius-Clapeyron, is that it only accounts for the change in bubble point as pressure changes. None of the theoretical methods take account of the change in relative volatility. Thus on increasing column pressure, even with the PCT held constant, there will be a decrease in separation and the purity of the products will worsen. This effect is secondary to the change in bubble point and will result in only a minor change but, if accuracy is required, *dT/dP* should be determined empirically, as described by Equations (12.67) and (12.91), or a more rigorous definition of *PCT* be developed from simulation.

The value of *dT/dP* increases as the bubble point of the product increases. We have seen that for LPG it is typically 3 to 4°C/bar. For a vacuum column operating at an absolute pressure of 30 mm Hg, with a tray temperature of 400°C, *dT/dP* will be in excess of 800°C/bar. Under these circumstances it is important that pressure changes are measured accurately. The pressure transmitter should be therefore located close to the temperature measurement so that changes in pressure drop across the column have no effect.

As mentioned earlier in this chapter, there are columns which do not lend themselves to tray temperature control, for example because temperature is insensitive to composition. Under these circumstances pressure compensation may be applied directly to what would otherwise be the MV of the temperature controller. The pressure compensation factor can be determined empirically. The approach is similar to that described for quantifying *dT/dP*. It is based on the assumption that an inferential can be developed based on pressure (P), the manipulated flow (F) and other independent variables. The manipulated flow may be reboiler duty, reflux, distillate or bottoms – depending on the choice of level control strategy.

$$Q = a_0 + a_1 P + a_2 F + \ldots\ldots \tag{12.96}$$

Differentiating

$$dQ = a_1 dP + a_2 dF \tag{12.97}$$

For no change in composition as pressure is changed

$$\frac{dF}{dP} = -\frac{a_1}{a_2} \tag{12.98}$$

Alternatively a linear (or nonlinear) function can be developed from process simulation.

12.11 Inferentials

While there are good reasons for locating tray temperature controllers away from the point at which the products are withdrawn, this does have a disadvantage. Because fractionation takes place between the temperature controller and the product draw, any change in the liquid and vapour traffic in this section of the column will change the correlation between product composition and tray temperature.

For example, if the temperature is controlled on tray 3, the bottoms composition will vary slightly as reflux is changed. This is illustrated in Figure 12.98. With no temperature control in place, reflux has a huge effect on bottoms composition. With tray 3 under temperature control the bottoms composition is kept close to target but not exactly so. The controller is effectively keeping constant the composition of the liquid on tray 3. As reflux is increased more separation takes places between tray 3 and the base of the column, thus improving the product purity.

By considering what else is taking place in the column, it is possible to develop an inferential which takes account of this effect. Figure 12.99 shows how the tray 3 temperature controller increases the reboiler duty to compensate for the cooling effect of the additional reflux. Figure 12.100 shows the effect on

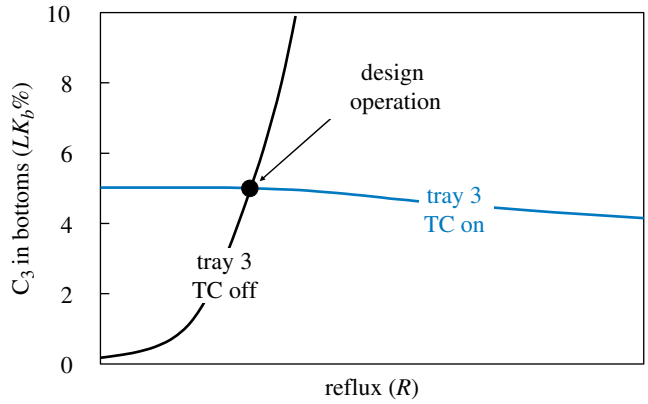

Figure 12.98 *Effect of reflux on bottoms quality*

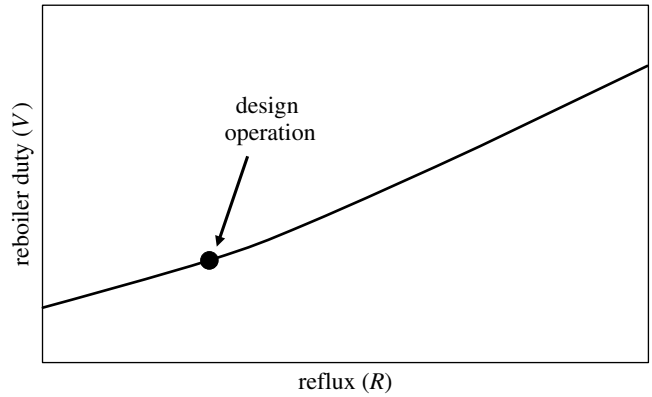

Figure 12.99 *Effect of reflux on reboiler duty*

distillate composition. With tray 3 under temperature control the bottoms purity remains approximately constant. The additional reflux and reboiler duty increase separation by improving the purity of the distillate. At the high reflux/reboil operation virtually all the C_4 is recovered from the distillate. As a result the bottoms yield increases as shown in Figure 12.101.

Figure 12.102 uses the data collected during this test to develop an inferential property calculation for bottoms composition. Despite the highly nonlinear behaviour displayed in the previous figures, a strong linear relationship exists between the C_3 content of bottoms and the reciprocal reboil ratio. This relationship can be expressed as

$$LK_b = a_0 + a_1 \frac{B}{V} \tag{12.99}$$

V can be any measure of reboiler duty. It might be the steam flow to a condensing reboiler. If heat integrated it might be the flow of heating fluid multiplied by its temperature change. For a fired reboiler it might be the fuel flow. Note that the flow B will usually be determined by the column level controller. For the inferential to respond quickly the level controller must be tightly tuned. If this is in conflict with a

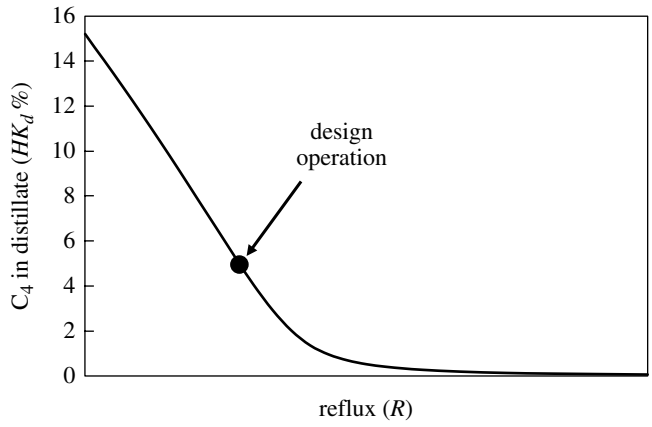

Figure 12.100 *Effect of reflux on distillate composition*

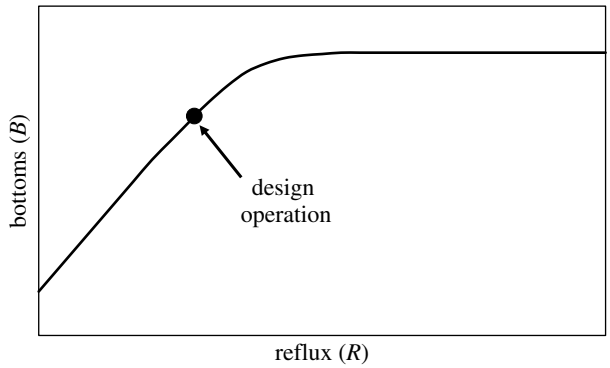

Figure 12.101 *Effect of reflux on bottoms flow*

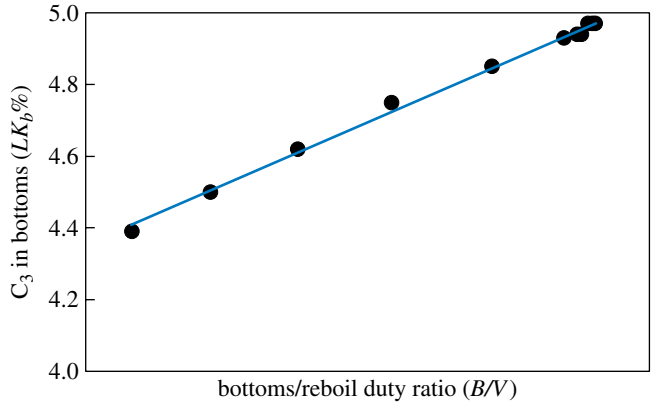

Figure 12.102 *Inferential for bottoms composition*

requirement to maintain, as steady as possible, the flow to the downstream unit then averaging level control should be retained. The effective bottoms flow is then determined by adding, to the measured flow, the rate of change of liquid volume in the column base. It is better, in the inferential calculation, to use the SP of the flow controllers rather than their PV. The argument for doing so is the same as that described for the pressure measurement used in a PCT. We do not want the inferential to reflect disturbances that will be handled by the basic controllers. While the PV of a flow controller will generally stay close to its SP, this is not true of slower controllers that might be included – such as those for temperature. Further a SP is usually subject to less noise than a PV.

There are a number of ways in which this inferential may be applied. At this stage, replacing the temperature controller with a virtual composition controller is not an option because the inferential is developed on the assumption that the tray temperature is held constant. It could be configured as a virtual analyser controller cascaded to the temperature controller. The approach needs some care since the input to the inferential is effectively the output of the temperature controller. If we use the actual reboiler duty, rather than its SP, as the input into the inferential, the lag introduced may make the cascade stable. Another approach would be to incorporate a correction term into the temperature measurement in much the same way as linear pressure compensation.

$$PV = T - \frac{dT}{dP}\left(P - P_{ref}\right) - K\left[\left(\frac{B}{V}\right) - \left(\frac{B}{V}\right)_{ref}\right] \tag{12.100}$$

K could be derived from historical process data, in much the same way as dT/dP was in Equation (12.67), or from simulation. Another option is to include the reciprocal reboil ratio with other variables, such as column pressure and tray temperature, in regression analysis to develop an inferential of the form

$$LK_b = a_0 + a_1 P + a_2 T + a_3\left(\frac{B}{V}\right) \tag{12.101}$$

This effectively incorporates *PCT* into the inferential. However, it might be preferable to retain this in the tray temperature controller – for example to permit its use if the inferential is out of service. Arithmetically, if a linear *PCT* is used, the two approaches are identical.

From Equation (12.67)

$$PCT = T + \frac{a_1}{a_2}\left(P - P_{ref}\right) \tag{12.102}$$

The inferential would then comprise

$$LK_b = b_0 + b_1 PCT + b_2\left(\frac{B}{V}\right) \tag{12.103}$$

Substituting for *PCT*

$$LK_b = b_0 + b_1\left[T + \frac{a_1}{a_2}\left(P - P_{ref}\right)\right] + b_2\left(\frac{B}{V}\right) \tag{12.104}$$

Equating coefficients with Equation (12.101)

$$b_0 = a_0 + a_1 P_{ref} \qquad b_1 = a_2 \qquad b_2 = a_3 \tag{12.105}$$

If a nonlinear *PCT* is required, e.g. because of large variations in pressure, the inferential in Equation (12.103) would instead be developed by regression based on this as one of the inputs. Any of the Antoine, Clausius-Clapeyron or Maxwell-Bonnell methods would then be considered. One approach would be to retain the *PCT* equation suggested by the chosen method and apply a least squares technique to adjust the coefficients to obtain the best predicted composition. Experience shows that doing so can result in coefficients very different from those derived theoretically and therefore make a significant improvement to the accuracy of the inferential. For example, if *PCT* is defined according to Equation (12.76) we can use regression to identify the coefficients (*b*) in the following inferential.

$$LK_b = b_0 + b_1 PCT + \ldots \tag{12.106}$$

$$\therefore LK_b = b_0 + b_1\left(T - B\left[\frac{1}{A - \ln\left(P\right)} - \frac{1}{A - \ln\left(P_{ref}\right)}\right]\right) + \ldots \tag{12.107}$$

To improve its accuracy we would use regression to identify the coefficients (*a*) in an inferential of the form

$$LK_b = a_0 + a_1 T + \frac{a_2}{a_3 - \ln\left(P\right)} + \ldots \tag{12.108}$$

By equating the coefficients of Equations (12.107) and (12.108), better estimates of the Antoine coefficients would then be given by

$$A = a_3 \qquad \text{and} \qquad B = -\frac{a_2}{a_1} \tag{12.109}$$

The coefficients in Equation (12.105) would then become

$$b_0 = a_0 - \frac{a_2}{a_3 - \ln\left(P_{ref}\right)} \qquad \text{and} \qquad b_1 = a_1 \tag{12.110}$$

Similarly, if *PCT* is defined according to Equation (12.79), we would derive by regression the inferential

$$LK_b = a_0 + \frac{a_1}{a_2 \ln\left[\frac{P}{P_{ref}}\right] + \frac{1}{T}} \ldots \tag{12.111}$$

The coefficients in Equation (12.106) would be unchanged. But a better estimate of the coefficient used in Equation (12.79) would be

$$\frac{R}{\lambda.MW} = a_2 \tag{12.112}$$

Another approach would be to perform a more general regression analysis to include nonlinear functions and not concern ourselves with converting it into a theoretical equivalent. As we have shown above, ln(*P*) should always be considered as an input and so inferentials of the form given by Equation (12.88) are likely to offer the best chance of success.

A very similar approach can be applied to the distillate composition. If the temperature is controlled on tray 18, the composition will change slightly as reboiler duty is changed. This is illustrated in Figure 12.103. With no temperature control in place, reboiler duty has a huge effect on distillate composition. With tray 18 under temperature control the distillate composition is kept close to target but not exactly so. The controller is effectively keeping constant the composition of the liquid on tray 18. As reboiler duty is increased more separation takes places between tray 18 and the top of the column, thus improving the product purity.

Again it is possible to develop an inferential which takes account of this effect. Figure 12.104 shows how the tray 18 temperature controller increases the reflux to compensate for the heating effect of the additional reboiler duty. Figure 12.105 shows the effect on bottoms composition. With tray 18 under temperature control the distillate purity remains approximately constant. The additional reboil duty and reflux

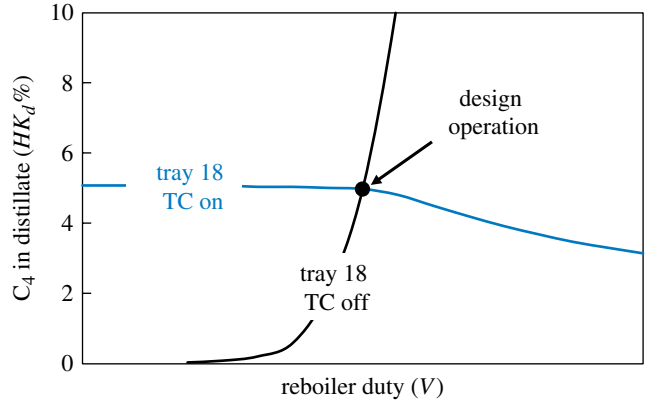

Figure 12.103 *Effect of reboiler duty on distillate quality*

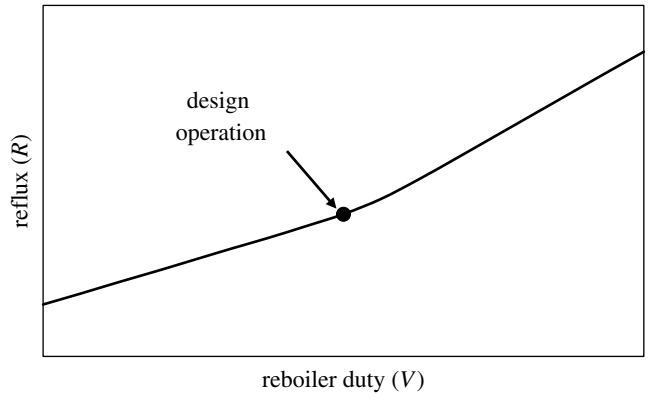

Figure 12.104 *Effect of reboiler duty on reflux*

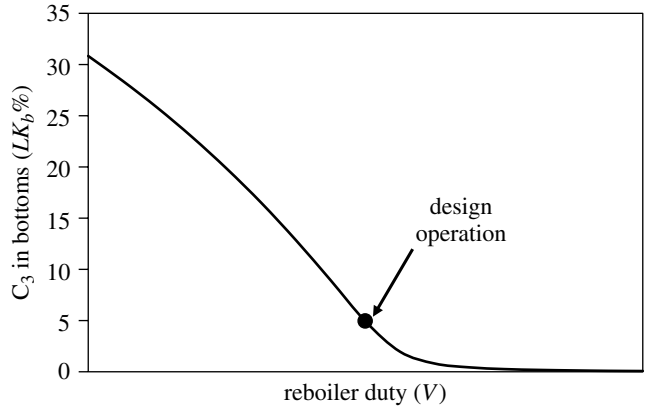

Figure 12.105 *Effect of reboiler duty on bottoms composition*

increase separation by improving the purity of the bottoms. At the high reboil/reflux operation virtually all the C_3 is recovered from the bottoms. As a result the distillate yield increases as shown in Figure 12.106.

Figure 12.107 uses the data collected during this test to develop an inferential property calculation for distillate composition. Again a strong linear relationship exists, this time between the C_4 content of distillate and the reciprocal reflux ratio. This relationship can be expressed as

$$HK_d = a_0 + a_1 \frac{D}{R} \tag{12.113}$$

If the material balance scheme is installed then R will be determined by the drum level controller. As described previously this scheme requires that the level controller be tightly tuned and so the inferential will respond quickly to disturbances. However, if the energy balance scheme is installed, the drum level controller will determine the flow D. Using tight level control may be in conflict with a requirement to maintain, as steady as possible, the flow to the downstream unit. Averaging level control can be retained by using the effective distillate flow – determined by adding, to the measured flow, the rate of change of liquid volume in the drum.

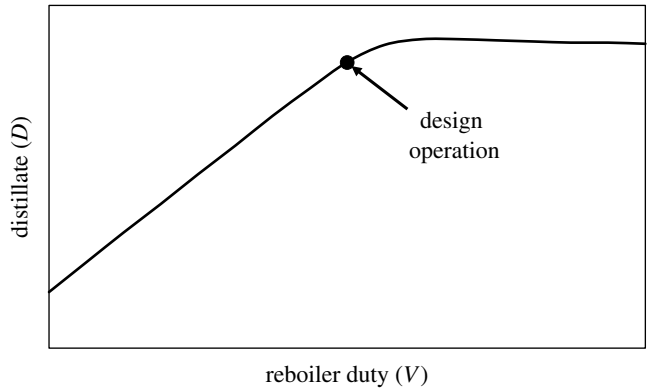

Figure 12.106 *Effect of reboiler duty on distillate flow*

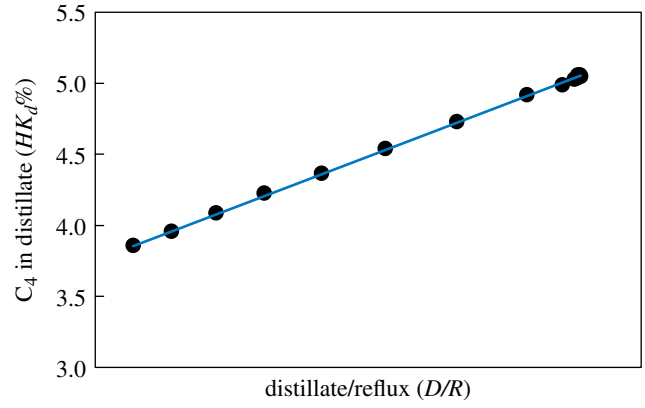

Figure 12.107 *Inferential for distillate composition*

The options for applying this correlation are equivalent to those for the bottoms composition controller. With care, it could be set up as composition controller cascaded to the temperature controller. Or a correction term could be incorporated into the temperature measurement.

$$PV = T - \frac{dT}{dP}\left(P - P_{ref}\right) - K\left[\left(\frac{D}{R}\right) - \left(\frac{D}{R}\right)_{ref}\right] \tag{12.114}$$

Or the reciprocal reflux ratio could be included with other variables, such as column pressure and tray temperature, in regression analysis to develop an inferential of the form

$$HK_d = a_0 + a_1 P + a_2 T + a_3\left(\frac{D}{R}\right) \tag{12.115}$$

Or the *PCT* could be retained and the distillate inferential could be based on this.

$$HK_d = b_0 + b_1 PCT + b_2\left(\frac{D}{R}\right) \tag{12.116}$$

As described in Chapter 9, care must be taken to include in the regression analysis only those process parameters that make engineering sense. But the form that those parameters take is not always immediately obvious. Figure 12.108 shows the true relationship between product composition and tray temperature at three operating pressures. The inclusion of only *P* and *T* in the inferential would imply that the relationship is a series of parallel straight lines as shown (dashed). Accuracy can be improved by including *compound* terms derived from independent variables – such as cross-products, ratios, etc. For example, the additional inclusion of *P.T* as an independent variable results in a closer match to the nonparallel lines, as shown in Figure 12.109. Alternatively, the use of a nonlinear function, e.g. $\log_{10}(HK_d)$, reduces the error by fitting curves rather than straight lines – as shown in Figure 12.110.

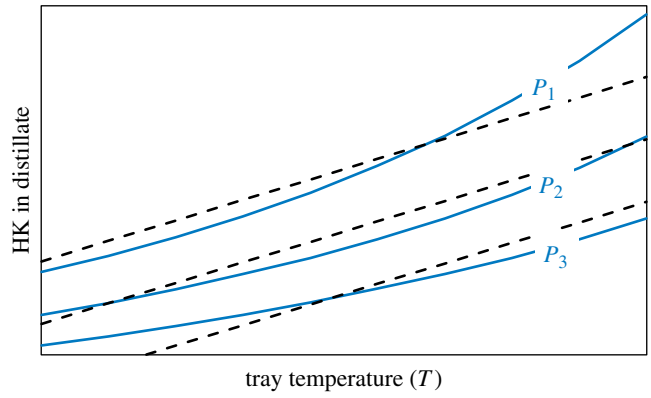

Figure 12.108 *Predicting composition from pressure and tray temperature (homogenic)*

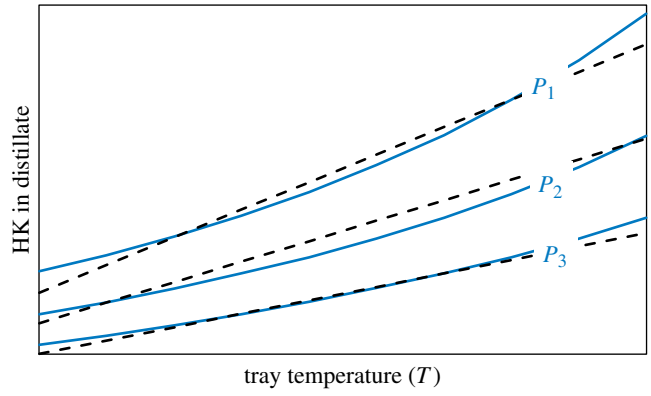

Figure 12.109 *Predicting composition from pressure and tray temperature (heterogenic)*

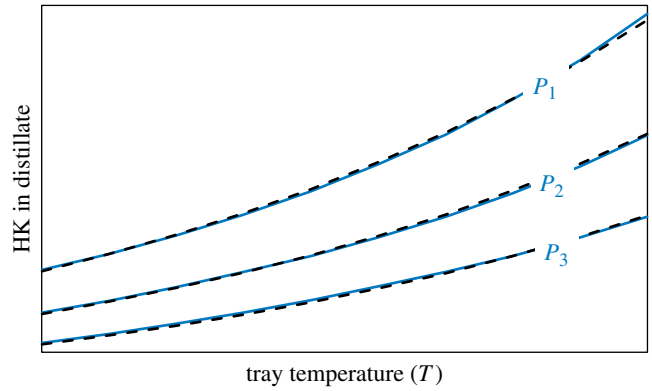

Figure 12.110 *Predicting composition from pressure and tray temperature (non-linear)*

Examination of the inferentials developed shows that they each include a measure of cut (the *PCT*) and fractionation (reflux or reboil ratio). While *PCT* is an effective measure of cut, there are a number of ways of including fractionation. For example Equation (12.117) shows part of an inferential developed using two tray temperature measurements where T_1 is higher up the column than T_2.

$$HK_d = a_0 + a_1T_1 - a_2T_2 \ldots\ldots \tag{12.117}$$

The coefficients a_1 and a_2 are positive numbers. Since we expect HK_d to increase with tray temperature we might be tempted to exclude T_2 and repeat the regression. However, this equation can be rewritten as

$$HK_d = a_0 + (a_1 - a_2)T_1 - a_2(T_2 - T_1) \ldots\ldots \tag{12.118}$$

Provided $(a_1 - a_2)$ is positive, the correlation makes sense by including a measure of cut (T_1) with a coefficient of the correct sign. The term $(T_2 - T_1)$ is a measure of the slope of the temperature profile and so of separation. Increasing separation will reduce HK_d and so the sign of a_2 is also correct. Indeed, it is possible on some columns to control product composition by controlling the difference between two tray temperatures by directly manipulating, for example, reboiler duty.

If pressure compensation is to be included it should not be necessary to apply it to both T_1 and T_2 since both values will be compensated by approximately the same amount. In Equation (12.118) T_1 could be replaced by its equivalent *PCT*. The inferential may exhibit unusual dynamic behaviour if the two trays are far apart. Step-testing will reveal if this is a problem. If so, it may be necessary to include an alternative measure of fractionation or omit it completely.

On high purity columns, expressing compositions as impurities, it is usually better to develop regressed inferentials for $\log(HK_d)$ and $\log(LK_b)$. We can approximate Equation (12.41) to

$$S \approx \frac{10000}{LK_b.HK_d} \tag{12.119}$$

So

$$\log(S) = 4 - \log(LK_b) - \log(HK_d) \tag{12.120}$$

We showed earlier in this chapter that separation is more linearly represented by log(S). The use of logarithms therefore helps accommodate the nonlinear relationship between composition and separation that is common on this type of column. It also has the advantage that the predicted values for HK_d and LK_b will never be negative.

A function (f) based on the bubble points of the light and heavy key components (T_{LK} and T_{HK}) would also be worth considering if other approaches fail. The bubble points (at normal operating pressure) are derived from the Antoine Equation. For the HK_d inferential T is the temperature on the tray selected in the upper section of the column.

$$HK_d = f\left(\frac{T - T_{LK}}{T_{HK} - T}\right) \tag{12.121}$$

For the LK_b inferential T is the temperature on the tray selected in the lower section of the column.

$$LK_b = f\left(\frac{T_{HK} - T}{T - T_{LK}}\right) \tag{12.122}$$

In addition to regressed inferentials there are a range of commercially available first-principle models. These use conventional heat and mass balances plus published correlations. Since much of this technology is proprietary, the content here is restricted to a summary of the published correlations.

12.12 First-Principle Inferentials

The Fenske Equation has already been covered as Equations (12.34) and (12.37). Knowing the relative volatility of the components and the target separation, the minimum number of theoretical stages (N_{min}) can be calculated.

Underwood's Method comprises two equations. The first includes all n components in the feed, where α is the volatility with respect to the least volatile component and x_f the mole fraction in the feed. Knowing the feed quality (q), the value of ϕ can be calculated.

$$1 - q = \sum_{i=1}^{n} \frac{\alpha_i \left(x_f\right)_i}{\alpha_i - \phi} \tag{12.123}$$

This value is then used to determine the minimum reflux ratio $(R/D)_{min}$. The equation includes the m components in the distillate, where x_d is the mole fraction in the distillate.

$$\left(\frac{R}{D}\right)_{min} + 1 = \sum_{i=1}^{m} \frac{\alpha_i \left(x_d\right)_i}{\alpha_i - \phi} \tag{12.124}$$

Gilliland's Correlation can then be used with the value of N_{min} from Fenske and $(R/D)_{min}$ from Underwood to determine the actual number of stages (N) from the actual reflux ratio (R/D).

$$\frac{N - N_{min}}{N + 1} = f\left(\frac{\left(\frac{R}{D}\right) - \left(\frac{R}{D}\right)_{min}}{\left(\frac{R}{D}\right) + 1}\right) \tag{12.125}$$

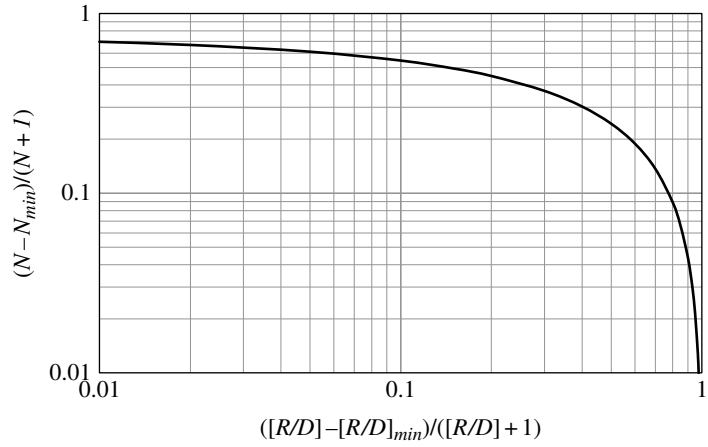

Figure 12.111 *Gilliland's correlation*

The correlation was originally published as a chart, as shown in Figure 12.111. A range of formulae have since been developed to describe it in equation form. The simplest of these is the *Eduljee Equation*.

$$\frac{N-N_{min}}{N+1}=0.75\left[1-\left(\frac{\left(\frac{R}{D}\right)-\left(\frac{R}{D}\right)_{min}}{\left(\frac{R}{D}\right)+1}\right)^{0.5668}\right] \tag{12.126}$$

Other equations include those developed by Chung and by Molkanov. Once N is known it can be used in a number of other correlations. For example the *Jafarey, Douglas and McAvoy Correlation* was developed by simplifying *Smoker's Equation* [31].

$$\log(S)=N\log\left[\alpha\sqrt{1-\frac{\frac{R}{D}+q}{\left(\frac{R}{D}+1\right)\left(\frac{R}{D}LK_f+q\right)}}\right] \tag{12.127}$$

Alternatively the *Colburn Equation* [32] gives the mole fraction of a chosen component in the vapour leaving the n^{th} tray (y_n) based on the mole fraction of the component in the bottoms (x_b).

$$\frac{y_n}{x_b}=\frac{\left(U^n-1\right)(\alpha-1)}{(U-1)}+1 \tag{12.128}$$

U is derived from the vapour-to-liquid molar flow ratio on the tray. For example, in a two product column, above the feed tray

$$U=\alpha\frac{R+D}{R} \tag{12.129}$$

Below the feed tray

$$U = \alpha \frac{V}{V + B} \tag{12.130}$$

A refinement is to add to $R + D$ the rate of change of reflux drum inventory. This helps overcome the problem caused by the process not yet reaching steady state. This is particularly useful if averaging level control has been installed on the drum in order to minimise flow disturbances to the downstream process. Similarly the rate of change of the liquid inventory in the base of the column can be added to $V + B$.

The number of trays (N) can also be determined empirically by adjusting it to fit sets of known conditions obtained from plant test runs. N need not be an integer and so can incorporate tray efficiency.

12.13 Feedforward on Feed Rate

Change in feed rate is a common source of process disturbance. We have seen that commissioning a single temperature controller (in the lower section of the column) manipulating fractionation may control bottoms composition well but makes control of distillate composition worse. The same effect applies to feedforward controllers. Figures 12.112 and 12.113 show the effect, with the energy scheme in place, of maintaining one of the MVs in proportion to feed rate but not the other. While keeping both in proportion to feed rate will result in virtually no change to product compositions, feeding forward to only one causes both product compositions to vary by substantially more than they would with no feedforward in place.

This is explained by consideration of the material balance around the upper section of the column as shown in Figure 12.7. Rearranging Equation (12.17) gives

$$\frac{D}{F} = \frac{V}{F} - \frac{R}{F} + 1 - q \tag{12.131}$$

As discussed in Section 12.4, the distillate cut (D/F) must be kept constant. Doing so, assuming no change in q, requires that V/F and R/F both be kept constant. Failure to do both will result in a change in cut and therefore a change in both product compositions.

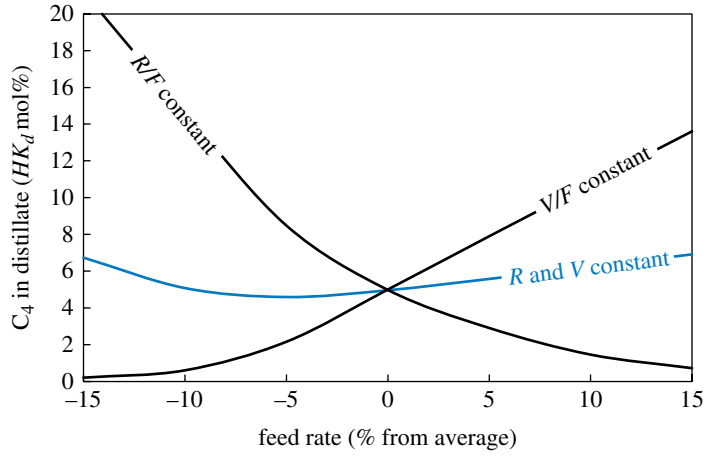

Figure 12.112 *Effect of feed rate feedforward on distillate (energy balance scheme)*

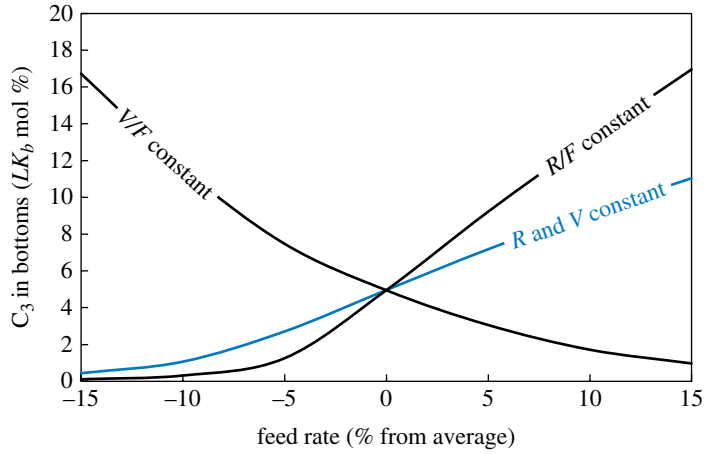

Figure 12.113 *Effect of feed rate feedforward on bottoms (energy balance scheme)*

Figures 12.114 and 12.115 show the same test with the preferred version of the material balance scheme in place. Again the distillate quality would be better controlled with no feedforward, rather than with just one of the ratios kept constant. While keeping either ratio constant marginally improves control of bottoms quality, keeping both constant will result in virtually no change.

While these figures show that both MVs need to change in proportion to feed rate this does not necessarily imply that this be by means of feedforward control. It would be quite common for one variable to be under tray temperature control and the other ratioed to feed rate. Indeed, this is similar to the configuration suggested by Ryskamp. However, ratio feedforward (with a feedback TC manipulating the ratio) offers three other potential benefits. Firstly, it gives a dynamic advantage. If tray temperatures respond slowly to changes in feed rate, then feedforward would permit tighter control of temperature. Secondly, it removes the need to re-tune the tray temperature controllers to compensate for the change in process gains caused by the change in feed rate. Finally, if tray temperature control is not practical or temporarily out of

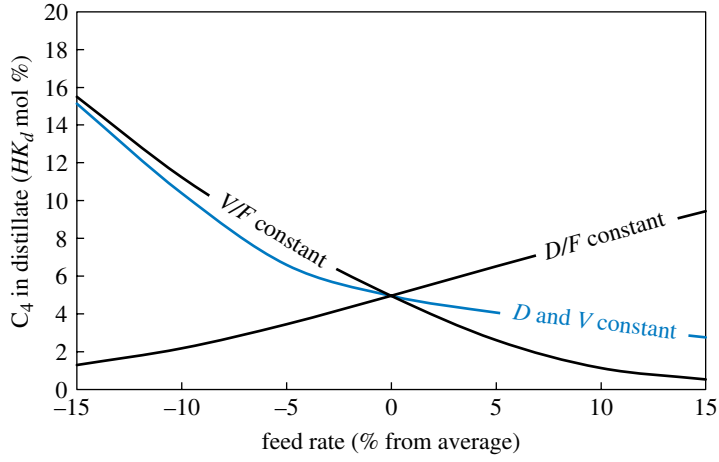

Figure 12.114 *Effect of feed rate feedforward on distillate (material balance scheme)*

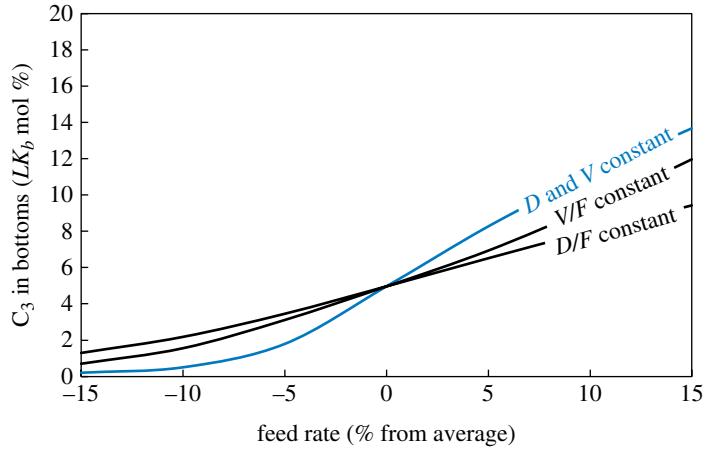

Figure 12.115 *Effect of feed rate feedforward on bottoms (material balance scheme)*

service, then feedforward would keep tray temperatures approximately constant. Figure 12.116 shows the application of full feed rate feedforward on a column with the energy balance scheme. Either or both of the reflux-to-feed and the steam-to-feed ratios can remain with operator entered SPs or their SP can be adjusted by a higher level control (such as tray temperature, inferential or on-stream analyser).

Dynamic compensation is likely to be necessary to ensure that the reflux and steam flows are adjusted at the right time. The method for tuning these deadtime/lead-lag algorithms is described in Chapter 6. Part of this procedure involves step-testing the DV, in this case feed rate, to obtain the dynamic response of the PV, in this case tray temperature. This can present a problem on some columns. Figure 12.117 shows the variation of some selected tray temperatures, in our 20 tray column, as feed rate is varied. The reboiler duty is fixed so, at very high feed rates, the increase in enthalpy it provides (per unit of feed) becomes insignificant and all tray temperatures approach the feed temperature and so no separation takes place. Conversely, at very low feed rates, all the feed is vaporised and ultimately its temperature reaches the bubble point of the heaviest component; so again no separation takes place. The slope of the appropriate line is the process gain between tray temperature (PV) and feed rate (DV), i.e. $(K_p)_d$. So, if we had chosen tray 18 for temperature control, we might experience difficulty obtaining the process dynamics – since the slope of this line changes sign. Since the value of $(K_p)_d$ is used in none of the tuning calculations in ratio feedforward this will not cause a problem with controller performance. However, if step-testing is performed in the region where the gain changes sign, model identification is likely to fail. While we are not concerned about the value of $(K_p)_d$, we do need values for θ_d and τ_d. It may be necessary to perform separate step-tests, each staying within a region where the slope is approximately constant, and then average the values obtained for deadtime and lag.

In the same way that we showed that two composition controllers interact, so will feedforward control. While it is correct at steady state to maintain a constant ratio between MVs and feed rate it is likely to be the case that the two MVs should not be changed at the same time. Indeed, our tuning of the dynamic compensation is designed to keep the product compositions constant. If, because of testing errors or the process being far from first order, we see transient disturbances in one of the compositions then it will be tempting to adjust the dynamic compensation by trial-and-error. But if we alter the timing of changes to reflux, for example, this will affect both product compositions – making it appear that the other feedforward compensation needs re-tuning. It is therefore unlikely that we would ever achieve perfect compensation and re-tuning should only be attempted if performance is particularly poor.

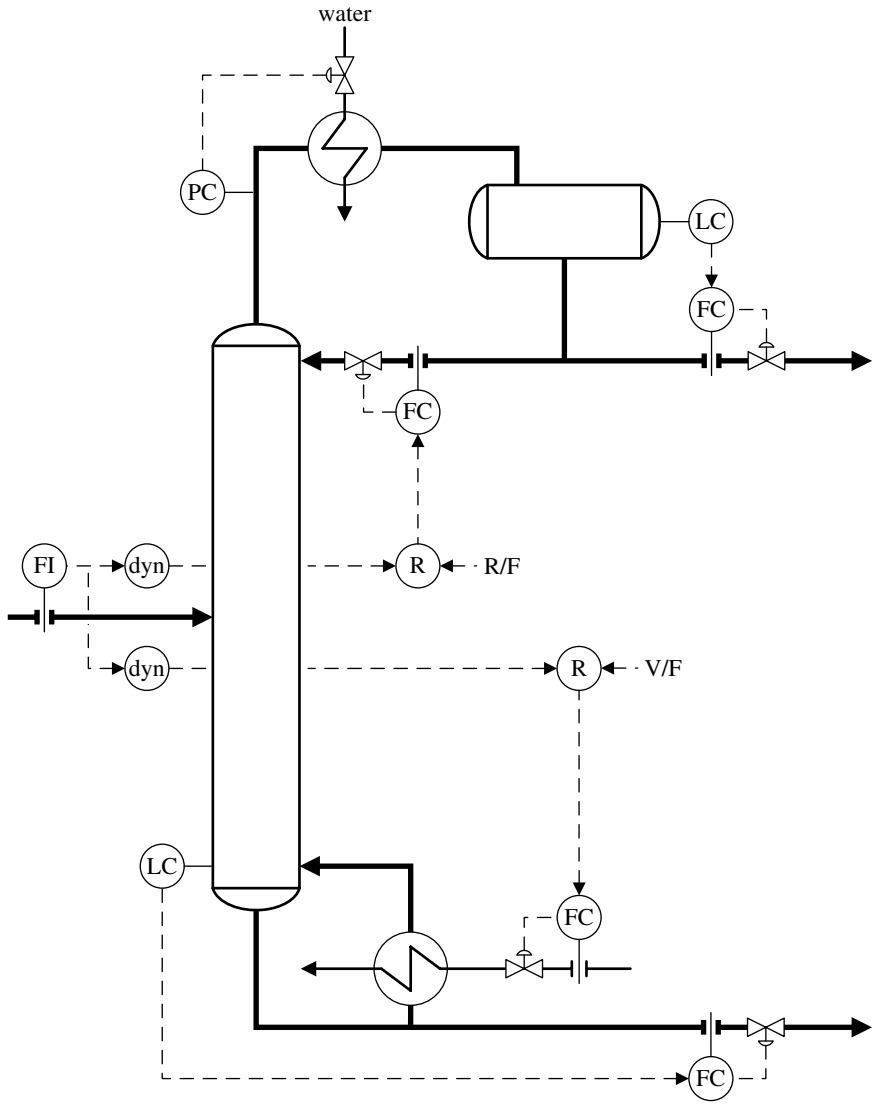

Figure 12.116 Feed rate feedforward (energy balance scheme)

12.14 Feed Composition Feedforward

Feedforward on feed composition can be a valuable enhancement but may not be practical. Firstly, it may require an on-stream analyser on feed. Few plant owners would install this as standard and there may not be sufficient economic justification to add it later. Secondly, it may not be possible to acquire an analyser that responds quickly enough. If the change in feed composition affects tray temperatures and/or inferentials before being reported by the analyser, then the feedback controller(s) will take corrective action. A delayed measurement of feed composition could then be less valuable than having no measurement.

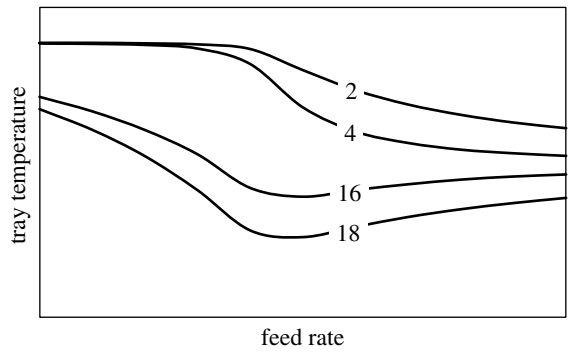

Figure 12.117 *Variation of tray temperatures with feed rate (energy balance scheme)*

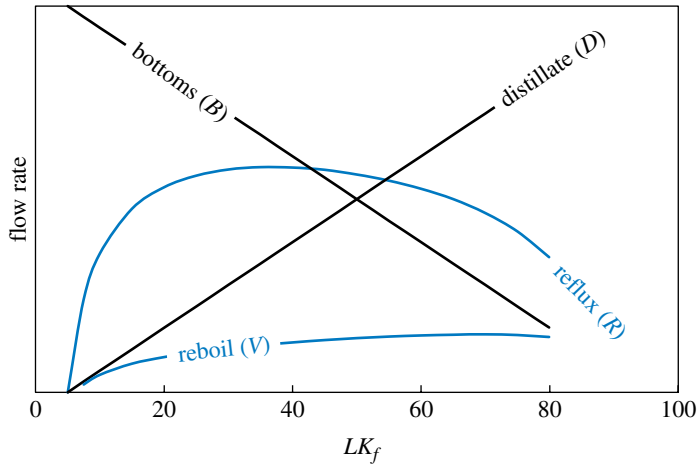

Figure 12.118 *Flow changes required as feed composition changes*

However, if the feed is produced by an upstream unit and routed directly to the column (i.e. not via storage), it may be possible to develop an inferential based on the operating conditions in that unit.

Another difficulty may be in the determination of the feedforward gain (K) that should be used. Unlike feed rate feedforward, feed composition feedforward requires a bias not a ratio algorithm and so K is not 1 (see Chapter 6 for explanation). Figure 12.118 shows how each of the possible MVs should be adjusted as feed composition changes. The shape of these lines was explained earlier in the chapter (see Figure 12.74). K is the slope of the line of the chosen MV. The required distillate and bottoms flows are linearly related to LK_f. So, if we have either version of the material balance scheme, then feedforward is straightforward. Differentiating Equation (12.25), for feedforward to distillate flow

$$K = \frac{dD}{d\left(LK_f\right)} = \frac{F}{LK_d - LK_b} \tag{12.132}$$

Or, if feedforward is to the bottoms flow

$$K = \frac{dB}{d\left(LK_f\right)} = \frac{-F}{LK_d - LK_b} \tag{12.133}$$

K is a function of feed rate (F). If feed rate is not constant this can be dealt with by adaptive tuning or feeding forward to the *D/F* (or *B/F)* ratio implemented for feed rate feedforward. The values of LK_d and LK_b are the product specifications. However, feeding forward to either distillate or bottoms flow, although beneficial, does not properly adjust fractionation. As Figure 12.118 shows, the slope of the reboiler duty and reflux lines change sign. If the average feed composition is close to this point, then it is unlikely that feedforward would be effective. Indeed, a better approach might be to assume that the line is horizontal and therefore exclude this part of the feedforward strategy. The situation is more complex if the energy balance scheme is installed because the reversal does not occur at the same feed composition. Care needs to be taken therefore that, for the range of possible feed compositions, it is reasonable to assume a linear relationship.

12.15 Feed Enthalpy Feedforward

Feed enthalpy feedforward is, rather fortunately, not often of great benefit. Changes in enthalpy are usually small compared to reboiler duty. If feedforward is justified then measuring enthalpy may present a problem. Provided the feed is below its bubble point and its composition is reasonably constant, it is sufficient to use temperature as the DV. If the feed is above its dew point, then the same approach is possible provided that also pressure is reasonably constant. If composition and/or pressure vary, provided they are measured, it is possible to calculate actual enthalpy. If the feed is partially vaporised a measurement of wetness is not possible. Since change in wetness represents a large change in enthalpy it is probably not realistic to make any assumption about its value. However, if the feed is partially vaporised by a feed preheater, it may be possible to calculate the change in enthalpy of the heating medium and add this to the feed enthalpy calculated upstream of the preheater. For example it is common for feed to be preheated by recovering heat from the bottoms product. It is likely that the composition of this stream is constant and its flow measured. Provided the inlet and outlet temperatures are known, then calculation of preheater duty is trivial. Indeed, some form of preheater duty control or feedforward technique may be desirable to prevent disturbances originating in the column being propagated to the column feed and aggravating the upset.

If we can successfully measure enthalpy we next need to determine what action to take when it changes. We could simply maintain the heat balance by adjusting the reboiler duty to compensate for the change. For this we need to be able to measure reboiler duty in consistent units. Further, since energy is entering the column at a different point, the liquid and vapour traffic in part of the column will change. So maintaining the heat balance is not sufficient to maintain product compositions.

Alternatively we could attempt to obtain the feedforward gains (K) empirically by plant testing, providing that we can introduce a disturbance into feed enthalpy. We may be able to determine K from analysis of historical data but, if collected while tray temperature (or some other composition) control was in service, it will only be possible to model steady state behaviour. Similarly we could identify K from steady state simulation. Dynamic compensation would then have to be tuned by trial-and-error.

12.16 Decoupling

We showed in Figures 12.25 and 12.26, with the material balance scheme in place, both product qualities are affected by both manipulated variables. Figures 12.119 and 12.120 show that this is also true with the energy scheme in place. So far we have considered schemes that control either the distillate composition or the bottoms composition but not both simultaneously. On most columns the two controllers would interact to the point of instability. There are a number of techniques available which help alleviate this problem.

We have seen that the Ryskamp scheme largely breaks the interaction in one direction so that corrections made to the bottoms composition have little impact on the distillate. Although the converse is not

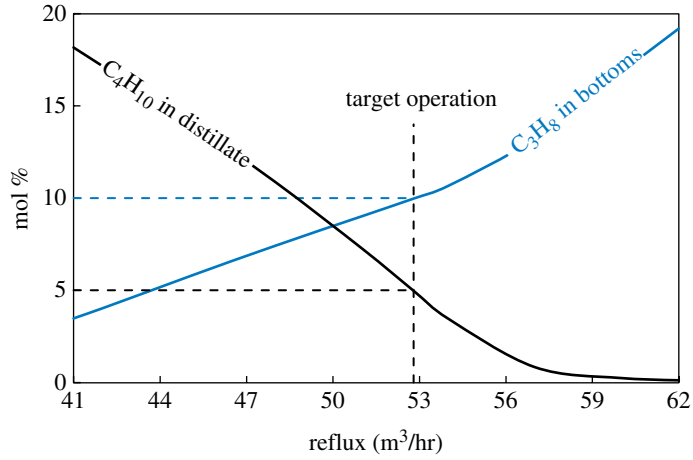

Figure 12.119 *Effect of reflux on product compositions*

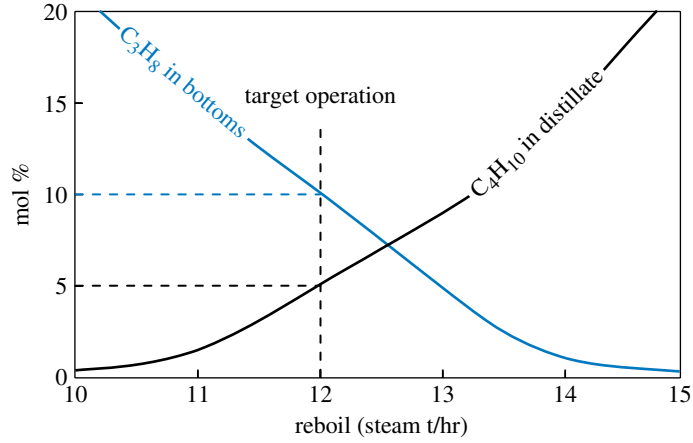

Figure 12.120 *Effect of reboil on product compositions*

true; an adjustment to the reflux ratio will affect the bottoms composition, but when its controller takes corrective action it will not 'fight' the distillate composition controller.

Another similar approach is to ratio, to feed rate, the MV not being used for control of composition. While not providing feedback control it does reduce variation in the uncontrolled composition and permits operator adjustment of the feedforward ratio target if necessary. If control of this composition is not important, then some variability might be a worthwhile price to pay to retain simplicity.

If dual composition control is required, then relative gain analysis, as described in Chapter 8, will help assess the level of interaction. While not an entirely accurate tool, because it only considers steady state interactions, it is indicative of the severity of the problem. Since, in any case, we need to perform step-tests to tune the composition controllers the additional effort involved in the analysis is minor. We can define, for the material balance scheme, PV1 and PV2 as respectively the distillate and bottoms product qualities, with MV1 and MV2 as the distillate flow and reboiler duty. Then, for example, from the slopes at the target operation in Figures 12.25 and 12.26, we can determine the process gain matrix

$$\begin{pmatrix} (K_p)_{11} & (K_p)_{12} \\ (K_p)_{21} & (K_p)_{22} \end{pmatrix} = \begin{pmatrix} 1.24 & -2.94 \\ -0.58 & -2.06 \end{pmatrix} \tag{12.134}$$

Switching to the energy balance scheme, MV1 becomes the reflux flow and, from Figures 12.119 and 12.120 we obtain the matrix.

$$\begin{pmatrix} (K_p)_{11} & (K_p)_{12} \\ (K_p)_{21} & (K_p)_{22} \end{pmatrix} = \begin{pmatrix} -1.27 & 3.94 \\ 0.55 & -5.06 \end{pmatrix} \tag{12.135}$$

Changing from the material balance to the energy balance scheme therefore changes the relative gain (λ_{11}) from 0.60 to 1.51. This change in the level of interaction might be a contributory factor in the decision as to which scheme to select.

Shinskey [33] describes a method for predicting these process gains, and hence relative gains, without the need for plant testing. By taking partial derivatives of Equations (12.25) and (12.41) it is possible to predict how product compositions vary with separation (at constant cut) and how they vary with cut (at constant separation). However, while it is possible on a real column to hold cut constant while varying separation, the converse is not readily achievable. It is not sufficient, when changing cut, to simply maintain a constant reboiler duty or (more precisely) reboil-to-feed ratio (*V/F*). Figure 12.121, based on the column described in Table 12.1, shows that this does not maintain a constant separation (*S*). As *D/F* is reduced HK_d will approach zero, LK_b will approach LK_f and so *S*, as defined by Equation (12.34), will become very large. Similarly as *D/F* is increased HK_d will approach HK_f, LK_b will approach zero and so *S* again will become very large. The process gain, predicted between product composition and cut, will thus not be the same as that determined by plant testing. Shinskey also suggests a linear separation model.

$$\frac{V}{F} = \beta \log(S) \tag{12.136}$$

This is required to convert predicted process gains based on directly manipulating separation to the more realistic situation where reboiler duty is the manipulated variable. The conversion factor (β) is determined from normal operating conditions. Figure 12.122 shows, for the same column, the true relationship between

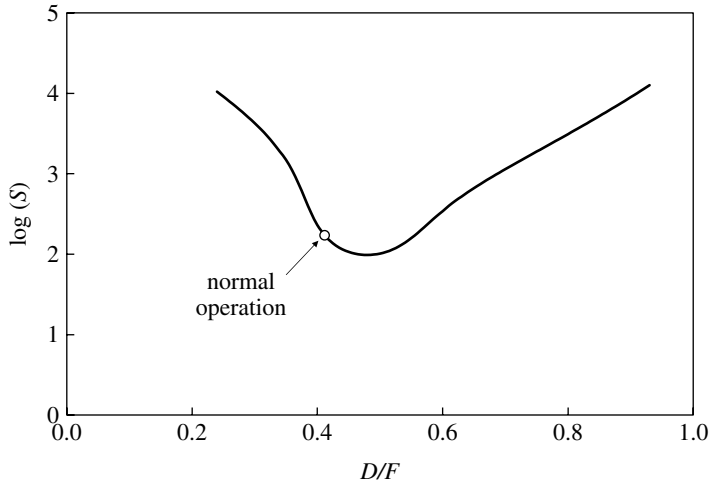

Figure 12.121 *Effect on separation of changing cut at constant reboiler duty*

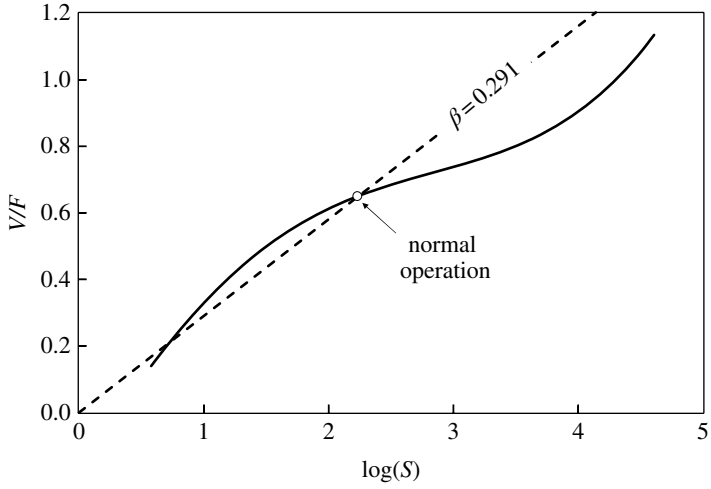

Figure 12.122 *Relationship between separation and reboiler duty*

the two variables versus the assumed linear relationship. The difference, between the value derived for β and the slope at normal operation, is significant and will introduce a substantial error in the prediction of process gains. As described in Chapter 8, estimates of relative gain can be very sensitive to small changes in process gains. It would therefore be unwise to base the selection of the material balance versus the energy balance scheme on predicted values and inadvisable to use them to design the composition controllers.

If one composition is considerably less important than the other, then it may possible to commission both controllers as normal PID controllers. The principle is to de-tune the less important controller so that, as it acts very slowly, the other faster controller can take corrective action to deal with any interaction. The more important controller should manipulate cut, as the dominant MV, or it may be impossible for it to reach its SP until the less important controller has set the cut to a feasible value. This may be in conflict

with dynamic considerations and result in bottoms quality being controlled by a MV at the top of the column and distillate quality being set by a MV at the bottom. The technique is to step-test as normal, to obtain the dynamics for the more important controller, but to tune and put this into automatic mode before step-testing for the other. When tuning the less important controller we would typically reduce the controller gain to around 25% of the design value – or even more as necessary for stability.

The ΣT/ΔT method is occasionally referred to texts on the subject of distillation control. It should only be considered on columns with the material balance control strategy. With this strategy in place a change in the distillate flow causes the column temperature profile to shift horizontally (see Figure 12.123). Because the profile largely maintains its shape there is no change in the difference between two tray temperatures. The sum, however, will change. A change in reboiler duty causes the profile to rotate (see Figure 12.124), thus the temperature in the upper section of the column will change in the opposite direction to one in the lower section. Thus the sum of the two temperatures will remain approximately constant. The difference, however, will change. In principle therefore ΣT responds only to changes in cut – i.e.

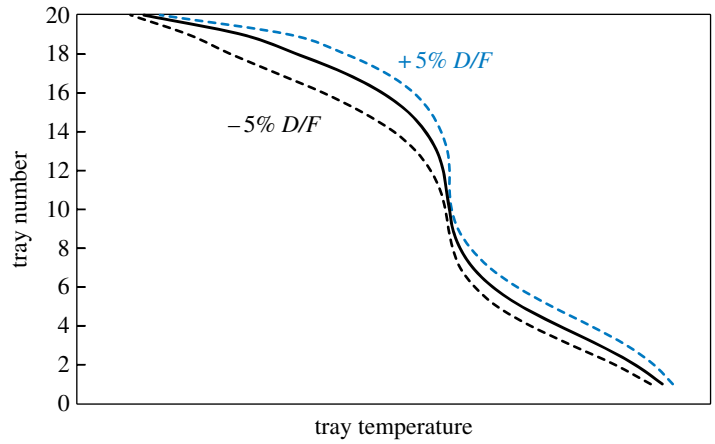

Figure 12.123 *Effect of changing cut (with material balance scheme in place)*

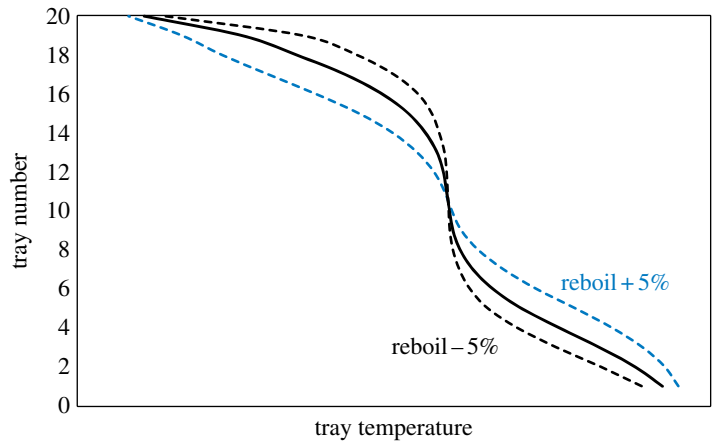

Figure 12.124 *Effect of changing reboil duty (with material balance scheme in place)*

distillate flow (or bottoms flow if the less preferred material balance scheme is in place). And ΔT responds only to changes in fractionation – i.e. reboil duty (or reflux).

Figure 12.125 shows how well the scheme would perform on our case study column. To keep the measurement in roughly the same range, ΣT is the average (half the sum) of the two temperatures. The principle appears to work well with very little change in ΣT over a wide range of reboiler duty and ΔT varying reasonably linearly. This would suggest that ΔT could be well controlled by manipulating reboiler duty.

Figure 12.126 shows the effect of varying distillate flow. Here the principle is not working well, with ΣT and ΔT both changing by similar amounts. Thus controlling ΣT by manipulating distillate flow will cause changes to ΔT. However, the tray temperatures were chosen to perform well as standalone controllers with the energy balance scheme in place. Selection based on the revised requirement would improve the decoupling. Performance might be further improved by the inclusion of coefficients such that ΣT is calculated as $(a_1T_1 + a_2T_2)$ and ΔT as $(b_1T_1 - b_2T_2)$. Nevertheless, the scheme as it stands has decoupled the controllers in one direction, which is sufficient to prevent the controllers fighting to the point of instability.

However, the decoupling takes no account of process dynamics. With the trays far apart one temperature will change at a different time to the other. Thus, when reboiler duty is changed, there will be a transient disturbance to ΣT. This will cause the controller to take corrective action and may trigger instability. It is

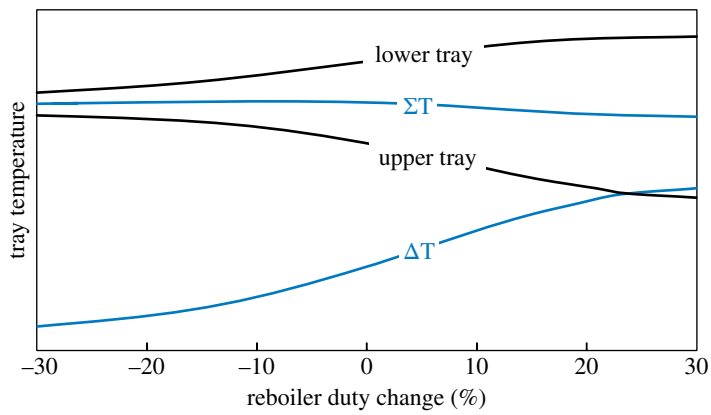

Figure 12.125 *Response of ΣT and ΔT to reboiler duty*

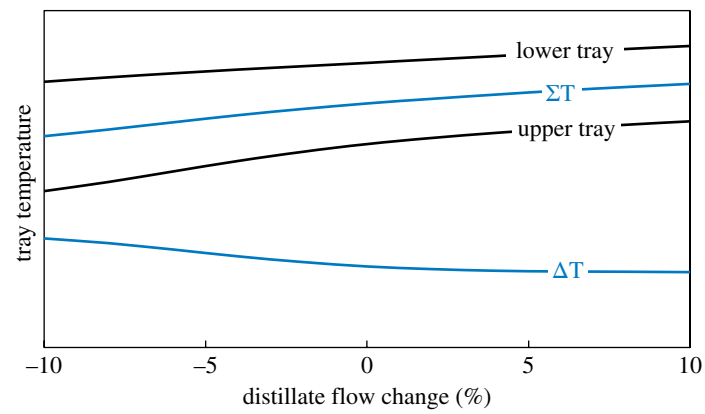

Figure 12.126 *Response of ΣT and ΔT to distillate flow*

not practical to dynamically compensate the temperatures since, for disturbances at the top of the column, the upper temperature will change first; for disturbances at the bottom of the column, the lower temperature will change first. The scheme should therefore only be considered if the dynamics of the two temperatures are similar – no matter what the source of the disturbance.

Success was claimed for a similar scheme based on on-stream analysers [34]. Here the PV controlled by reboiler duty was defined as the average of HK_d and LK_b while that controlled by distillate flow was defined as $(HK_d - LK_b)$. It is likely that analyser deadtime was large compared to that of the process so that any difference in dynamics between the top and bottom of the column would be, as a proportion of the overall delay, much smaller.

In Chapter 8 a more rigorous approach to decoupling is described. Configured using the DCS block structure it is a complex implementation and prone to a number of problems. For the reasons given it should only be considered if, for some reason, a MPC package cannot be applied.

12.17 Multivariable Control

One of the common applications of MPC is providing dual composition control on distillation columns. It is often the most practical way of resolving interactions. Here we work through a simple example of its design. Figure 12.127 shows the addition of MPC to a column with an energy balance scheme.

Notable by their omission are the tray temperature controllers. Some implementers believe this to be the better approach. Since two tray temperature controllers will interact they argue that they should be decoupled within MPC. But often the starting point for MPC implementation is a working temperature controller on one tray. Others argue that, if a basic controller is working well, it should be retained. It helps linearise the process and provides graceful degradation when MPC is out of service. A third view is that the temperature controller should be retained as a back-up scheme but this means the operator needs training in both the normal control configuration and the back-up scheme. There is no universally correct approach; each should be considered on a per case basis.

In our example the two CVs are both compositions, either inferentials or analysers. Our MVs are reflux and reboil. Plant testing gave the process gain matrix shown in Table 12.10. MPC will therefore predict the process behaviour

$$CV_1 = -0.962MV_1 + 4.17MV_2 + bias_1 \tag{12.137}$$

$$CV_2 = 0.806MV_1 - 5.32MV_2 + bias_2 \tag{12.138}$$

MPC continuously updates the bias terms as required so that the predicted CV matches the actual CV. It does this using the full dynamic model whereas our example is written on a steady-state basis. If we assume that when both CVs are at their targets of 5%, the reflux is 56.5 and the reboiler steam rate 14. We can therefore rewrite Equations (12.137) and (12.138) as

$$-0.962MV_1 + 4.17MV_2 = 4.027 \tag{12.139}$$

$$0.806MV_1 - 5.32MV_2 = -28.941 \tag{12.140}$$

Figure 12.128 shows a plot of these lines of constant composition. Figure 12.129 shows the operating constraints, to which has been added a maximum reboiler steam flow (of 16). Figure 12.130 shows the feasible operating area. Most MPC packages include a linear program (LP) for optimisation. This technology can only find the most profitable *node*, where constraints cross. There may be a more profitable

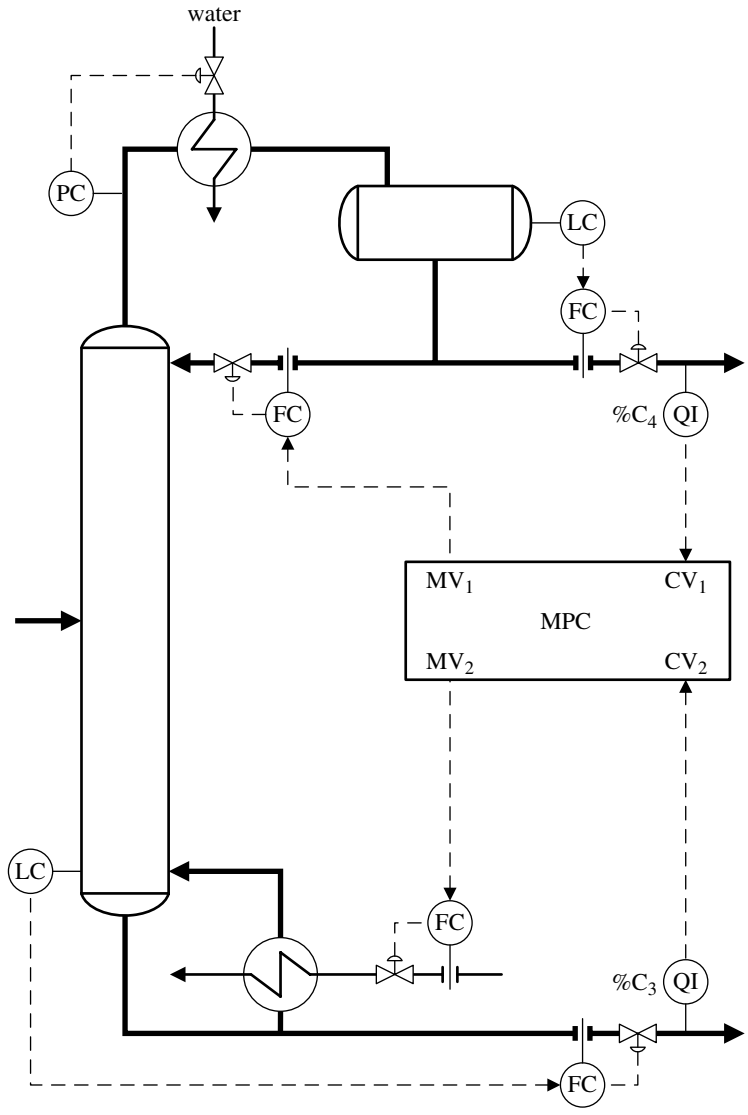

Figure 12.127 *MPC on column with energy balance scheme*

Table 12.10 *Process gain matrix*

		MV_1 Reflux	MV_2 Reboil
CV_1	% C_4 in distillate	−0.962	4.17
CV_2	% C_3 in bottoms	0.806	−5.32

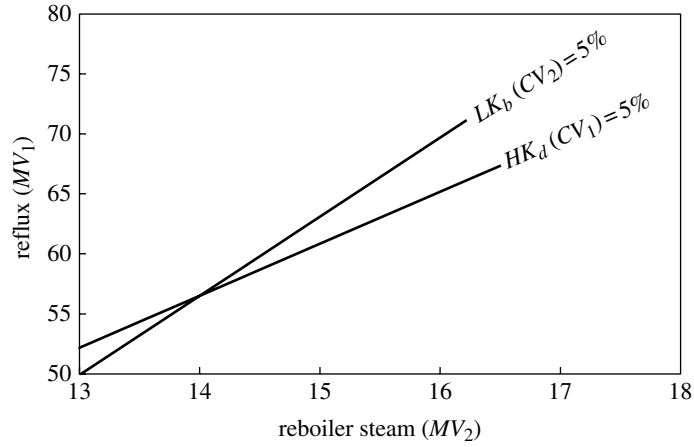

Figure 12.128 *Contours of constant composition*

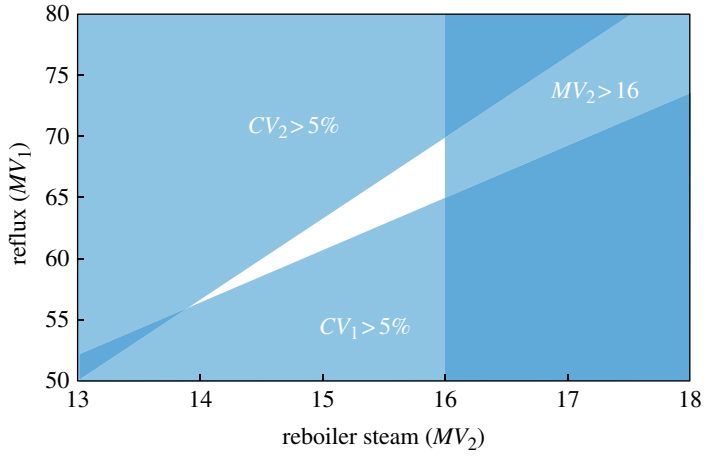

Figure 12.129 *Operating constraints*

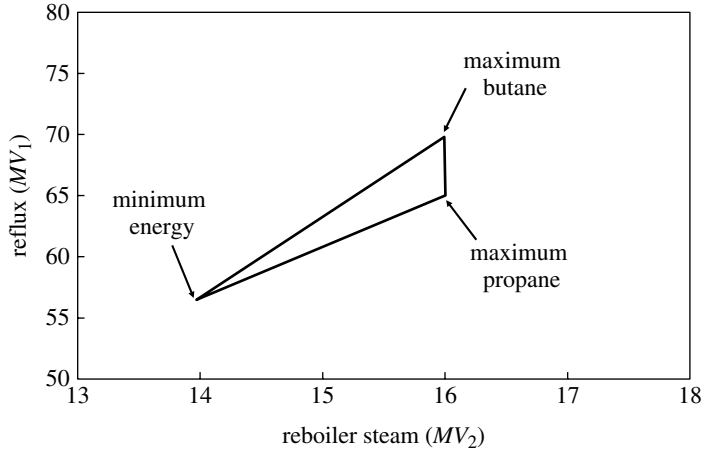

Figure 12.130 *Feasible operating region*

operating point within the feasible operating area. A nonlinear optimisation technique would be required to identify this point; this is covered later in this section.

So MPC will select one of the three nodes:

- the 'minimum energy' case, where both products are exactly at their specifications
- the 'maximum propane' case, where the propane is at specification and the maximum permitted reboiler duty is used to recover as much C_3 as possible from bottoms
- the 'maximum butane' case, where the butane is at specification and the maximum permitted reboiler duty is used to recover as much C_4 as possible from distillate

What drives MPC to select a node are the *objective coefficients* (or *cost coefficients*). These are applied to each MV (Q) and, in most MPC packages, also to each CV (P). If m is the number of CVs and n the number of MVs, then the MPC *objective function* or *cost function (C)* is given by

$$C = \sum_{i=1}^{m} P_i CV_i + \sum_{j=1}^{n} Q_j MV_j \tag{12.141}$$

As we shall show later, there are several advantages to using real economics in the controller. However, with our controller in its current form, this is not immediately obvious. What economic value would be placed on reflux, or on product compositions? To facilitate this we include additional variables in the controller. In this case we include the two products flows as additional CVs. No constraints are placed on these variables so they do not affect the feasible operating region. The product flows simply provide a more convenient way of defining the controller economics. Table 12.11 shows the extended gain matrix.

Note that, since MPC minimises a cost function rather than maximising a profit function, variables generating a revenue have negative cost coefficients. In our example P_3 and P_4 (for the product flows) will be negative and Q_2 (for reboiler steam) will be positive.

In adopting this approach it is important that the gain matrix is consistent with the rules of mass balance. In our example, at constant feed rate, any change in the yield of one product must be reflected by an equal and opposite change in the yield of the other. If this were not the case the controller would 'believe' it could increase the total product yield by making more of one and less of the other. This would clearly be 'perceived' as very attractive economically and would result in the controller adopting an operating strategy that almost certainly would be wrong. Thus, in this case, $(K_p)_{31}$ must be equal and opposite to $(K_p)_{41}$ and $(K_p)_{32}$ must be equal and opposite to $(K_p)_{42}$. Since plant testing is unlikely to produce exactly this result some adjustment will be necessary. Some commercial MPC model identification software packages permit rules of mass balance to be defined and will identify the best model fit that satisfies these rules. With others it is manual exercise. Similarly if feed rate were included as an additional MV, as shown in

Table 12.11 *Extended process gain matrix*

		MV_1 Reflux	MV_2 Reboil
CV_1	% C_4 in distillate	−0.962	4.17
CV_2	% C_3 in bottoms	0.806	−5.32
CV_3	propane flow	−1.000	5.40
CV_4	butane flow	1.000	−5.40

Table 12.12, mass balance requires that $(K_p)_{33}$ and $(K_p)_{43}$ sum to 1. If the material balance scheme were implemented on the column, then the MVs would be distillate flow and reboiler steam. Plant testing gave the results shown in Table 12.13. In this case, for the mass balance to be correct, $(K_p)_{31}$ must be −1, $(K_p)_{32}$ must be 0 and $(K_p)_{33}$ must be 1. There is no need to include propane flow as a CV since its inclusion as an MV already permits an objective coefficient to be assigned.

If the MPC package does not permit CVs to be given objective coefficients then the coefficients for the MVs should be modified to take account of the effect that changing each MV has on each CV.

$$Q'_j = Q_j + \sum_{i=1}^{m} \left(K_p\right)_{ij} P_i \tag{12.142}$$

Using the gains in Table 12.11, if we consider first the 'maximum butane' case, we see from Figure 12.130 that at this point the reboil steam is 16 and the reflux 69.7. Thus, in moving from the 'minimum energy' case, the change in butane production is given by

$$\Delta CV_4 = 1.000 \times \left(69.7 - 56.5\right) - 5.40\left(16 - 14\right) = 2.4 \tag{12.143}$$

This is matched by an equal loss of propane production. For this change to be economically correct the value of the additional butane must be greater than the cost of the additional steam.

$$-2.4P_3 + 2.4P_4 < 2Q_2 \qquad \text{or} \qquad P_3 - P_4 > 0.833Q_2 \tag{12.144}$$

Similarly, for the 'maximum propane' case, the reboil steam is 16 and the reflux 65.2. Thus,

$$\Delta CV_3 = -1.000 \times \left(65.2 - 56.5\right) + 5.40\left(16 - 14\right) = 2.1 \tag{12.145}$$

Table 12.12 *Inclusion of feed rate as a MV*

		MV_1 Reflux	MV_2 Reboil	MV_3 Feed rate
CV_1	% C_4 in distillate	−0.962	4.17	−0.098
CV_2	% C_3 in bottoms	0.806	−5.32	−0.446
CV_3	propane flow	−1.000	5.40	0.310
CV_4	butane flow	1.000	−5.40	0.690

Table 12.13 *Extended process gain matrix (with material balance scheme)*

		MV_1 Distillate	MV_2 Reboil	MV_3 Feed rate
CV_1	% C_4 in distillate	1.045	−2.17	0.253
CV_2	% C_3 in bottoms	−0.760	−2.17	−0.647
CV_3	butane flow	−1.000	0	1.000

For this to be profitable

$$2.1P_3 - 2.1P_4 < 2Q_2 \qquad \text{or} \qquad P_4 - P_3 > 0.952Q_2 \tag{12.146}$$

The purpose of this analysis is to demonstrate the impact of process economics on MPC. In this case, if the price difference between the two products approaches or exceeds the unit cost of steam, it would be incorrect to operate at the 'minimum energy' point. It is common practice to treat objective coefficients as weighting factors that are adjusted, often by trial-and-error, to force the controller to drive the process to what is believed to be the optimum operation. Indeed, some engineers will apply Equation (12.142), even if the MPC package supports assigning objective coefficients to CVs – arguing that this makes it easier to adjust coefficients to force a particular strategy. Adjusting coefficients in this way risks better achieving the wrong objective and so losing money when MPC is commissioned.

If real process economics are used then, during the testing phase, it will become apparent that the operating strategy suggested by MPC is different from the established strategy. We also show, later in this section, that this can be often be established from historical data collected before step-testing is even begun. Rather than simply adjusting the objective coefficients, the difference should be reconciled. Either MPC does not comprise a faithful model of the process, e.g. because the process gains are incorrect or a key model is missing, or the choice of economics is wrong. If none of these problems exist then the inevitable conclusion is that the current operating strategy is wrong. Given the complexity of many processes this should not be unexpected. It is likely that never before has such a detailed analysis of process optimisation been undertaken.

There are several other advantages to the use of true process economics. It is common for process operators to artificially constrain the MVs. In our example, consider the impact of the operator placing an upper limit on reflux of 60. As Figure 12.131 shows, this would severely constrain the maximum butane or maximum propane operating modes. With the use of real economics it is possible to quantify the cost of this.

For the 'maximum butane' case, Figure 12.131 shows the steam rate reduced to 14.5. The change in butane yield caused by constraining the reflux is given by

$$\Delta CV_4 = 1.000 \times (60.0 - 69.7) - 5.40(14.5 - 16) = -1.6 \tag{12.147}$$

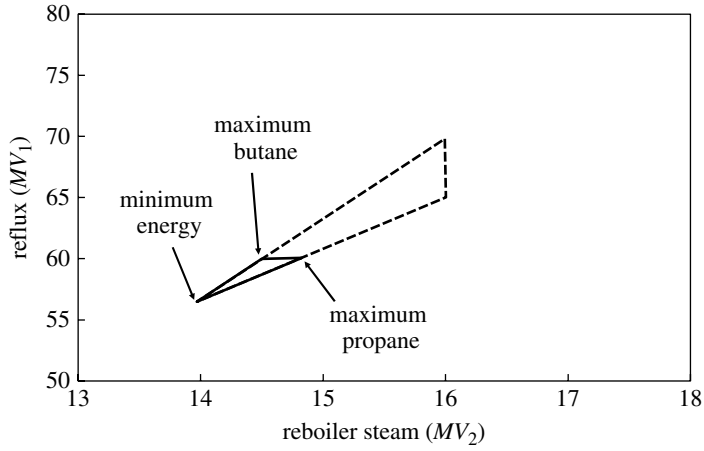

Figure 12.131 Over-constrained reflux

Taking account of the steam saving, the change in the value of the cost function is

$$\Delta C = -1.6 \times \left(P_4 - P_3\right) - 1.5Q_2 \tag{12.148}$$

Expressed per unit change in the MV

$$\frac{\Delta C}{\Delta MV} = \frac{-1.6 \times \left(P_4 - P_3\right) - 1.5Q_2}{69.5 - 60.0} \tag{12.149}$$

This value is known as the *reduced cost* or *reduced gradient*. By convention this is reported as a negative number if constrained by a high limit and positive if at a low limit. The amount by which the MV can change before another constraint is reached is the *allowable increase*. Some MPC packages make these values available to the engineer. If not, then a spreadsheet package will permit the engineer to build a steady state simulation of MPC. Most spreadsheet packages then automatically generate this information.

Performing a similar calculation for the 'maximum propane' case, Figure 12.131 shows the steam reduced to 14.8. The change in propane yield is therefore

$$\Delta CV_3 = -1.000 \times \left(60.0 - 65.2\right) + 5.40\left(14.8 - 16\right) = -1.3 \tag{12.150}$$

Taking account of the steam saving, the change in the value of the cost function is

$$\Delta C = -1.3 \times \left(P_3 - P_4\right) - 1.2Q_2 \tag{12.151}$$

A further advantage of the use of real economics is that they readily provide the economic incentive for process debottlenecking projects. If we assume that the constraint placed on reboiler duty is real then, if we want to maximise the yield of one of the products, there is benefit in relaxing this constraint. For example relaxing it to 17 would result in expanding the feasible area as shown in Figure 12.132. For the 'maximum butane' case, Figure 12.132 shows the reflux increased to 76.3. The change in butane yield resulting from increasing the reboiler duty is given by

$$\Delta CV_4 = 1.000 \times \left(76.3 - 69.7\right) - 5.40\left(17 - 16\right) = 1.2 \tag{12.152}$$

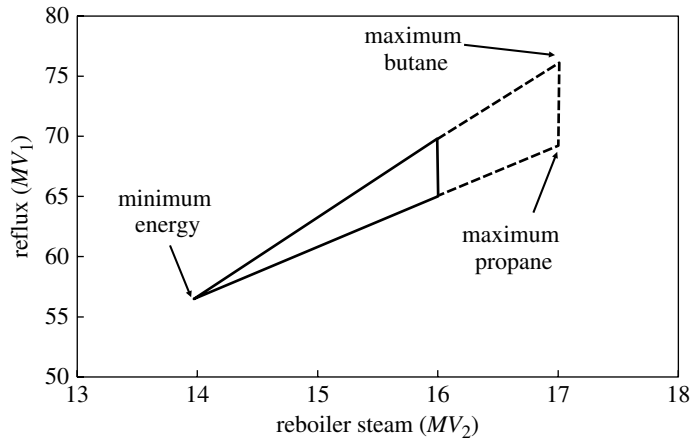

Figure 12.132 *Relaxing the reboiler constraint*

Taking account of the steam cost, the change in the value of the cost function is

$$\Delta C = Q_2 - 1.2 \times \left(P_4 - P_3 \right) \tag{12.153}$$

Repeating the calculation for the 'maximum propane' case, Figure 12.132 shows the reflux increased to 69.5. The change in propane yield caused is given by

$$\Delta CV_3 = -1.000 \times \left(69.5 - 65.2 \right) + 5.40 \left(17 - 16 \right) = 1.1 \tag{12.154}$$

Taking account of the steam cost, the change in the value of the cost function is

$$\Delta C = Q_2 - 1.1 \times \left(P_3 - P_4 \right) \tag{12.155}$$

We can also use MPC to quantify the value of changing constraints on CVs. As an example, let us assume that the price of butane and propane are identical but the opportunity exists to produce a higher purity propane (containing no more than 1% C_4), which can be sold at a higher price. Knowing the normal operation, with both products at their 5% specifications, we can calculate the biases in Equations (12.137) and (12.138)

$$CV_1 = -0.962 MV_1 + 4.17 MV_2 + 0.973 \tag{12.156}$$

$$CV_2 = 0.806 MV_1 - 5.32 MV_2 + 33.941 \tag{12.157}$$

Solving these equations, if CV_1 is reduced to 1%, shows that reflux is increased to 68.6 and reboiler steam to 15.8. The change in propane yield is therefore

$$\Delta CV_3 = -1.000 \times \left(68.6 - 56.5 \right) + 5.40 \left(15.8 - 14 \right) = -2.2 \tag{12.158}$$

Taking account of the steam cost, the change in the value of the cost function is

$$\Delta C = 1.8 Q_2 - 2.2 \times \left(P_3 - P_4 \right) \tag{12.159}$$

Since P_3 and P_4 are currently the same this new operating mode is more costly and needs to be justified by the higher propane price. Expressed per unit change in CV_3, Equation (12.160) gives the *shadow price* or *Lagrange multiplier*.

$$\frac{\Delta C}{\Delta CV_3} = \frac{1.8 Q_2 - 2.2 \times \left(P_3 - P_4 \right)}{2.2} \tag{12.160}$$

To make it attractive to produce, the price of the higher purity propane must exceed that of normal propane by at least this value. If the increase is more than this amount, then the 'maximum propane' mode would become more attractive and should therefore be analysed to determine the actual increase in profit.

Finally, the use of real economics means the MPC objective function is a real measure of process profitability. This can then be used to assess the value of the application. While the minute-to-minute value is probably too noisy and subject to change if the objective coefficients are changed, it can be used offline. For example, the mean C_4 content of propane before implementation might have been 4.1%. With MPC in place it was increased to 4.5% and so the improvement is 0.4%. We repeat this calculation for the other limiting constraints such as the C_3 content of bottoms and the reboiler steam flow.

We then run a simulation of MPC first with the normal constraints and then with the constraints tightened by the improvement. For example the constraint on C_4 in propane is moved from 5% to 4.6%. The difference in objective function between the two simulation runs will be profit improvement (per hour, if flows are measured per hour).

While using real economics is an ideal, there are sometimes practical difficulties in achieving this. For example, it is common to include column pressure as a MV. The resulting gain matrix is shown in Table 12.14.

The relationship between product composition and pressure is relatively straightforward to identify. Indeed, this is done in developing the PCT. But the impact that pressure has on product yields is often difficult to quantify during step tests. As a result $(K_p)_{33}$ and $(K_p)_{43}$ are often omitted. The controller therefore will only manipulate pressure to relieve a constraint; it sees no economic incentive to adjust it otherwise. The solution often adopted is to define an objective coefficient for the pressure MV, usually a small positive value in the belief that pressure should be minimised to reduce energy requirements. However, this may result in an opportunity to significantly improve profitability being overlooked. For example, increasing pressure would alleviate a condenser constraint, permitting increased recovery of the more valuable product. A better approach would be to find some other means of identifying the missing process gains, e.g. by regression of historical data or from process simulation.

At the beginning of this section we considered the retention of the tray temperature controller. Assuming this is in the lower section of the column and manipulates reboiler steam, then the SP of this controller becomes a MV, instead of that of the steam flow controller. However, we may wish to include steam flow as a CV so that we can apply an objective coefficient. The gain matrix then becomes that shown in Table 12.15.

Table 12.14 *Inclusion of pressure as a MV*

		MV_1 Reflux	MV_2 Reboil	MV_3 Pressure
CV_1	% C_4 in distillate	−0.962	4.17	−1.43
CV_2	% C_3 in bottoms	0.806	−5.32	2.78
CV_3	propane flow	−1.000	5.40	−2.30
CV_4	butane flow	1.000	−5.40	2.30

Table 12.15 *Retention of tray TC*

		MV_1 Reflux	MV_2 Tray TC
CV_1	% C_4 in distillate	−0.600	0.418
CV_2	% C_3 in bottoms	−0.006	−0.750
CV_3	propane flow	−0.330	0.650
CV_4	butane flow	0.330	−0.650
CV_5	steam flow	0.130	0.125

With the controller configured in this way, when the reflux was step-tested the tray TC took corrective action to maintain the temperature at SP. This affects the response of all five CVs. Thus if, once MPC is operating, the TC is switched to manual (for example because of a measurement problem) the process gains will revert to those in Table 12.12. Of particular concern would be $(K_p)_{21}$ since its value changes sign. It is unlikely that MPC could continue to operate reliably with such an error. MV_2 would have to be configured as a *critical* MV so that MPC is automatically disabled if the TC is switched to manual.

12.18 On-Stream Analysers

Much of the application of on-stream analysers has been covered in general in Chapter 9. Here we focus on those issues specific to their use on distillation columns.

One of our objectives is to minimise the sample delay; this can be achieved by locating the analyser as far up-stream as possible. For example, it is common to withdraw the distillate sample from the discharge of the product pump. But consideration should be given to withdrawing it from the overhead vapour line. This avoids the delay caused by the condenser and reflux drum. Plus the velocity of the vapour in the sample line will be far greater than liquid velocity. Some care should be taken with design of the sample system. It would be unwise to simply tee into the overhead line. Vapour can begin to condense on the internal surface of the line and any sample taken could then contain liquid containing more HK than the distillate. A better approach is to insert a probe into the top section of the column, through the elbow in the overhead line. The sample line to the analyser should be heated and insulated to ensure no condensation takes place. The vapour sample would only be representative of distillate composition if the condenser is total. However, if a vapour product is only produced intermittently, then it may be practical to detect production and temporarily disable the use of the analyser measurement in any control scheme.

On trains of distillation columns there may be other possible methods of reducing sample delay. Consider the arrangement in Figure 12.133 in which our case study LPG splitter is the last column. It is common to have specifications on the C_2 and C_4 content of propane. We have considered schemes that might control the latter by manipulation of the reflux on the LPG splitter. But to control the C_2 content we manipulate the reboiler duty on the preceding de-ethaniser. A composition controller based on a C_2 analyser on the propane product will respond very slowly to changes in its MV. A better approach would be to locate an analyser on the de-ethaniser bottoms. The C_2/C_3 ratio will be close to the C_2 content of propane but provides a much earlier indication and hence much faster disturbance rejection. Similarly it is common to have C_3 and C_5 specifications for butane. Control of C_3 would likely be by manipulation of the LPG splitter reboiler duty. However, control of C_5 would be manipulation of the reflux on the debutaniser – two columns upstream. Measuring the C_5/C_4 ratio of debutaniser distillate would provide a huge dynamic advantage. Of course costs need to be taken into account, but the incremental cost of an additional analyser, if installed in the same housing, is much smaller than the cost of a standalone installation. And further use may be made of such analysers – for example as part of feedforward strategies.

12.19 Towers with Sidestreams

The logic followed in designing control strategies for two-product columns can be extended to those with one or more sidestreams. The pressure controller, as before, is the first designed – using much the same approach as that described. The next step, the level control strategy, depends on the process configuration.

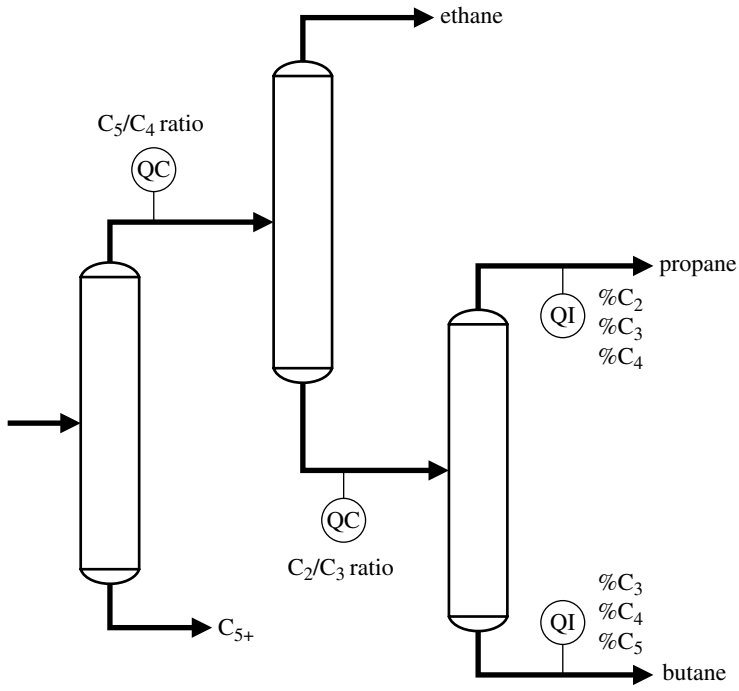

Figure 12.133 *Reducing analyser sample delay in distillation trains*

Figure 12.134 shows a column with no liquid distillate. There is little choice but control reflux drum level with the reflux. Assuming feed rate is not available as a MV, column level may be controlled conventionally by manipulating bottoms flow as shown. If the bottoms flow is too small to provide sufficient control then manipulation of the sidestream flow is an option.

With three liquid products we can elect to reserve the sidestream flow as a MV for composition control. There are three options, each of which is simply the addition of the sidestream to schemes we have already described. Figure 12.135 shows its addition to the energy balance scheme. Figure 12.136 shows its addition to the preferred material balance scheme. Figure 12.137 shows its addition to the less preferred material balance scheme.

The column LC can, if the more usual MVs cannot be used for some reason, be configured to manipulate the sidestream flow. Figure 12.138 is the result of making this change to the scheme shown in Figure 12.135. By adjusting the sidestream, the internal reflux is changed and so also the liquid that accumulates in the base of the column.

We have to give some additional consideration to the composition controllers. By adding the sidestream we have increased by one the number of MVs – i.e. the sidestream draw flow. But we potentially add two composition targets. We add a third key component – *middle key (MK)*. We have our existing composition targets now expressed as %MK in distillate and %MK in bottoms, but we may also have specifications for both %LK and %HK in the sidestream. Since we now have more PVs than MVs, we cannot meet all four product composition specifications. One of compositions will always be in giveaway (better than specification). If it is always the same one, then we can safely ignore it and design a scheme to control the remaining three. However, if the operation and/or the targets cause the three limiting specifications to change, then all four must be included in the controller and some logic added to automatically select which will be controlled. This function is readily provided by MPC packages.

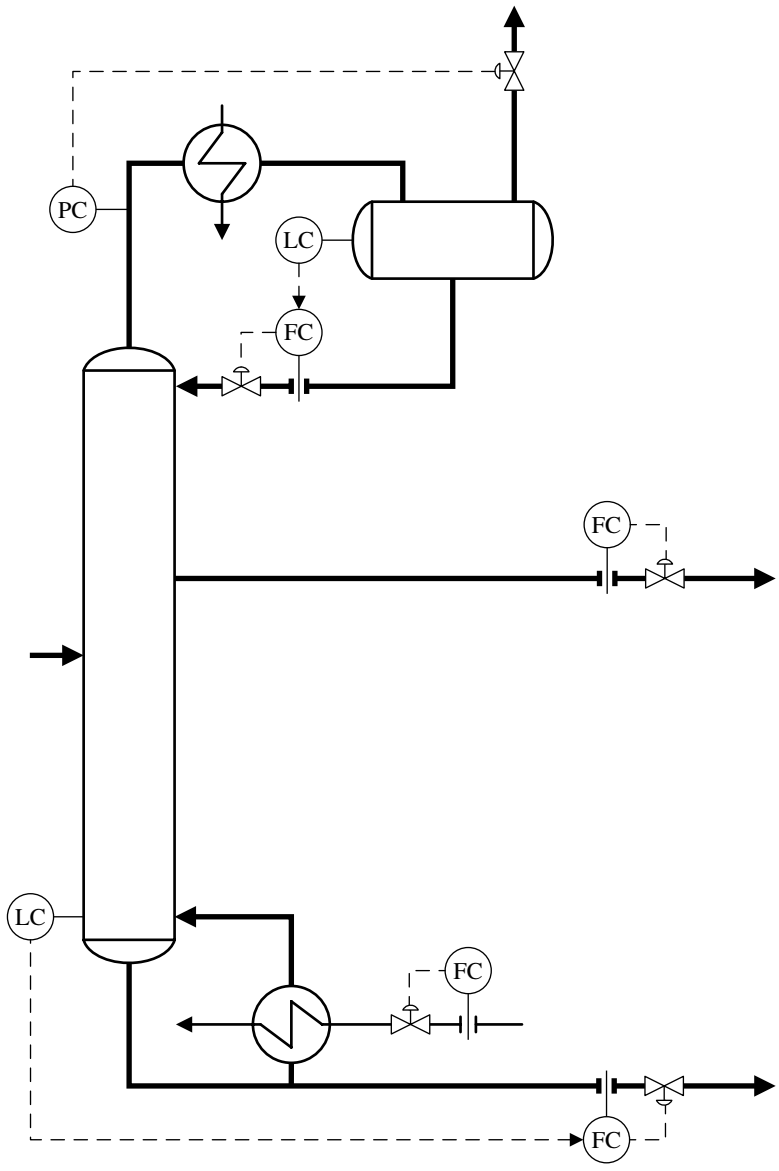

Figure 12.134 Reflux drum level control with no liquid distillate

12.20 Column Optimisation

Once the column has schemes which provide effective energy and material balance and composition control there may still remain a number of variables which can be manipulated to improve profitability. It may be possible to adjust feed rate, feed composition or feed enthalpy. There is usually scope to adjust column pressure. And, if there is a large difference in product prices, compositions can be adjusted to be better than specification.

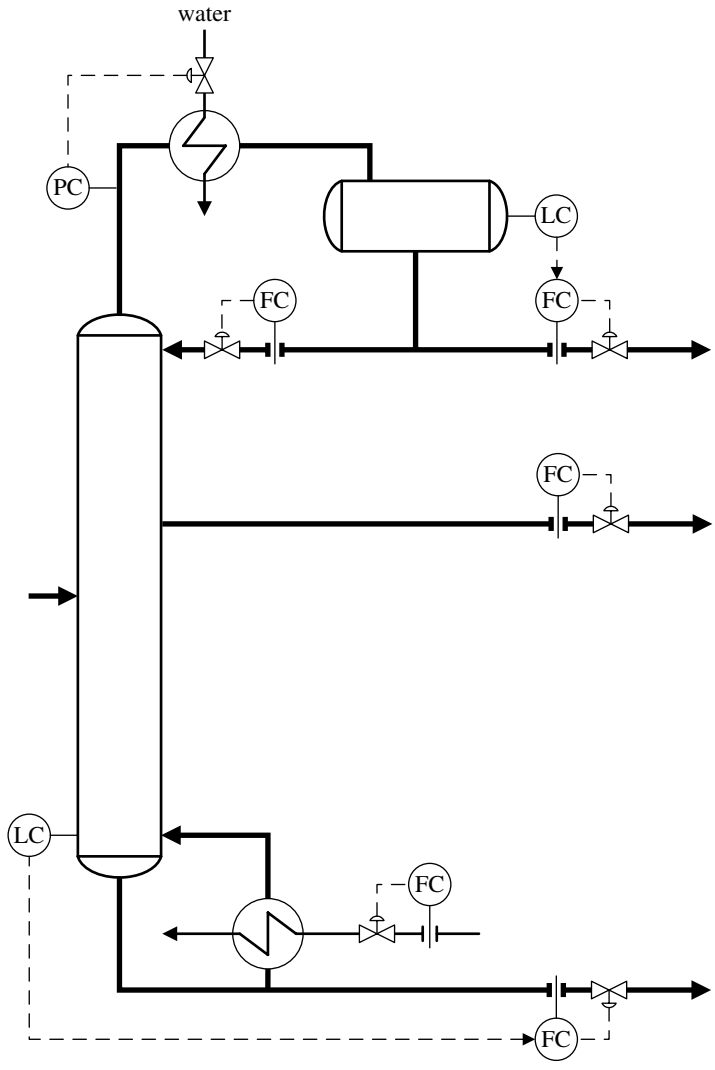

Figure 12.135 *Sidestream added to energy balance scheme*

In adjusting the operation several of a wide range of equipment constraints may be encountered. These include condenser duty which may limit because of high coolant inlet temperature (e.g. air-fins in hot weather) or because there is a maximum permitted coolant exit temperature (e.g. corrosion by salt water). The condenser limit might be approached because the column pressure is too low such that the dew point at the top of the column approaches the coolant temperature. High feed enthalpy can similarly overload the condenser, as can fouling on either the tube or shell side.

Similarly the reboiler may constrain. If duty is provided by heat integration, then a low heating medium inlet temperature may not give sufficient LMTD. Hot oil based heating systems are subject to a minimum outlet temperature, below which high viscosity gives pumping problems. Too high a column pressure will increase the bubble point so that again the LMTD is insufficient. On fired reboilers a maximum fuel rate may be reached, constrained by combustion air availability, tube skin temperature, burner pressure, etc.

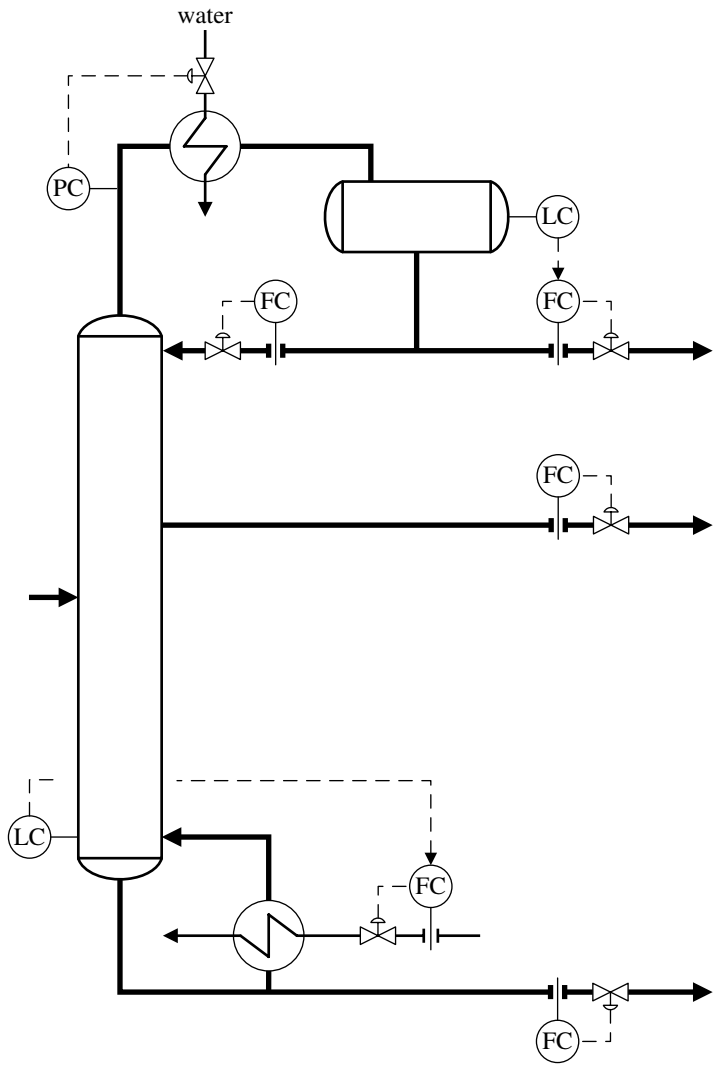

Figure 12.136 *Sidestream added to preferred material balance scheme*

There may be metallurgical constraints on temperature. And, like condensers, reboilers can be subject to fouling.

Parts of the unit may have hydraulic limits. If there is a vapour product, its flow rate may be limited by the pipework or, if present, its compressor. Restrictions may apply to pumps on feed, products or reflux. They can also apply to the flow of coolant through the condenser and heating fluid through the reboiler. As we saw at the beginning of this chapter the column may blow, weep or flood.

In selecting the optimisation technology, we need to determine what form the problem takes. If the number of available MVs exceeds the number of active constraints, then there is at least one degree of freedom and some form of nonlinear real-time optimiser will be required. If there are no degrees of freedom, then MPC is sufficient.

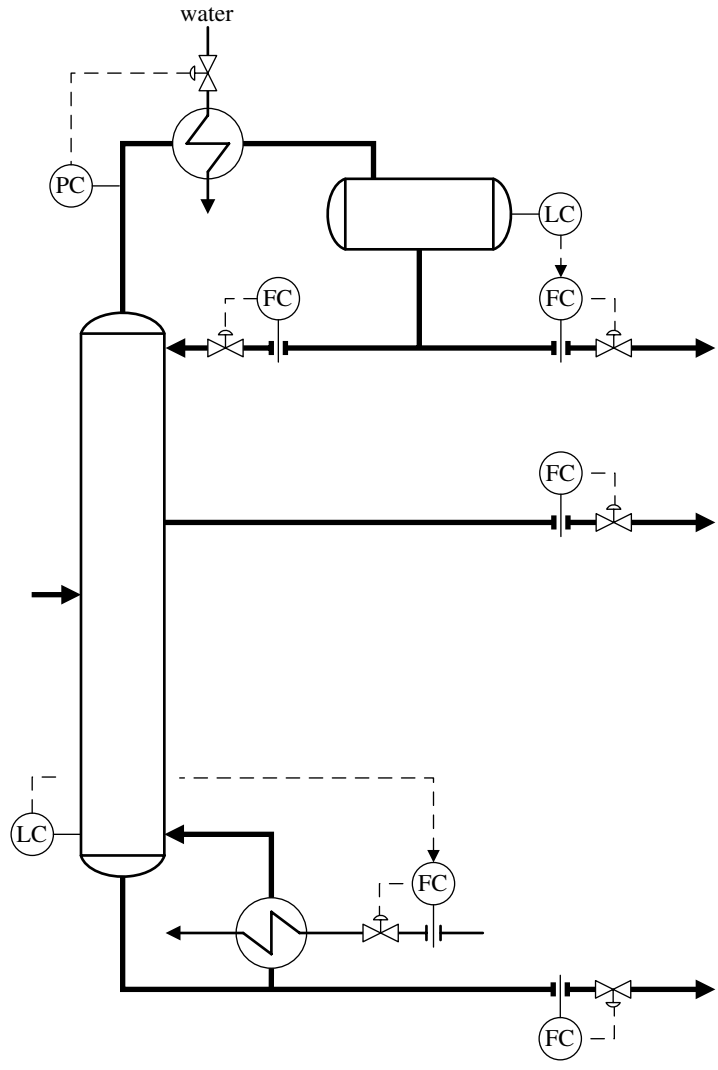

Figure 12.137 Sidestream added to less preferred material balance scheme

12.21 Optimisation of Column Pressure

We have seen that reducing pressure reduces the temperature at which the column operates and hence improves relative volatility. Figure 12.139 shows the effect this has on separation (at constant cut). If, instead of keeping the reboiler duty constant as we reduce pressure, we manipulate the duty to keep the product compositions constant, then we get the result shown in Figure 12.140. While designing the column to operate at minimum pressure would minimise its energy requirements, construction cost must also be considered. Figure 12.141 shows that, as pressure is reduced, the condenser inlet temperature falls. This could require a greater condenser area and hence larger equipment plus the increased cost of the support structure. It could also result in the inlet temperature falling below that of the available cooling fluid and

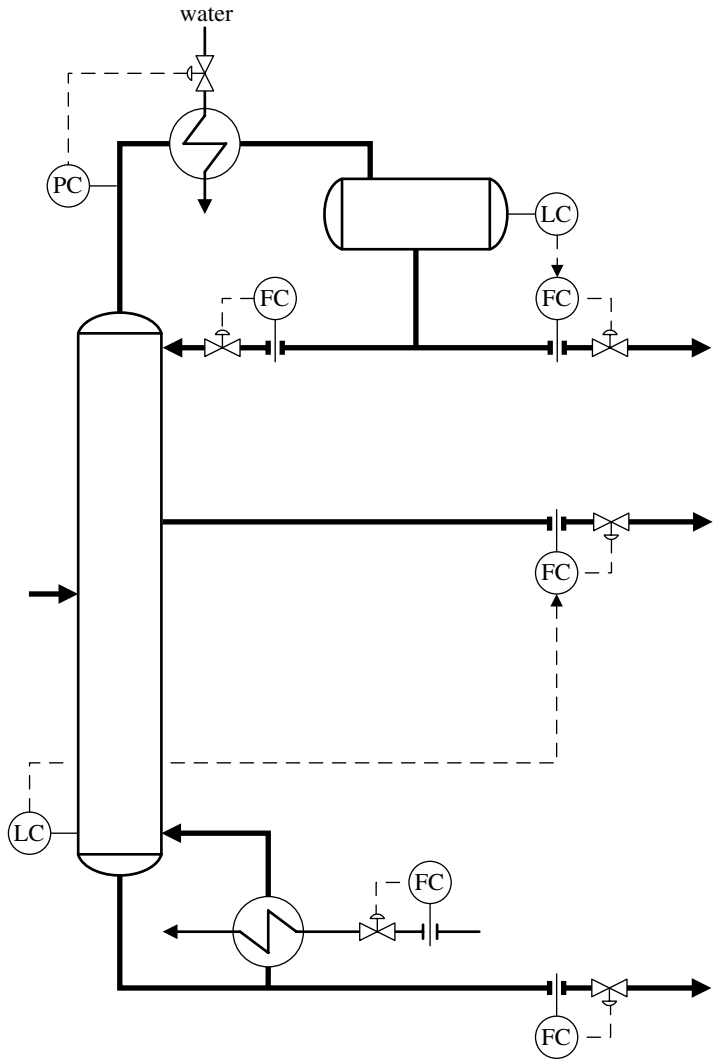

Figure 12.138 *Sidestream used for column level control*

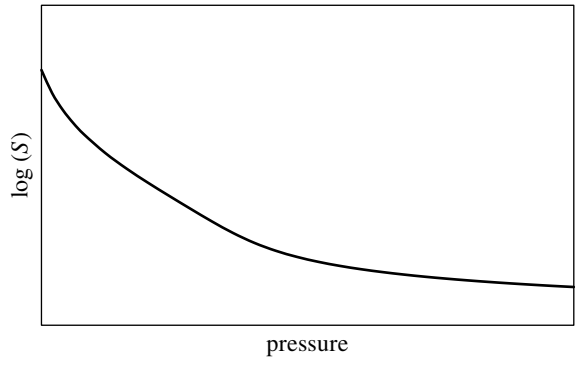

Figure 12.139 *Effect of pressure on separation*

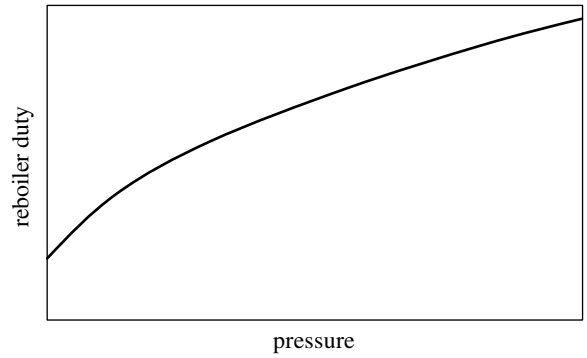

Figure 12.140 *Effect of pressure on reboiler duty*

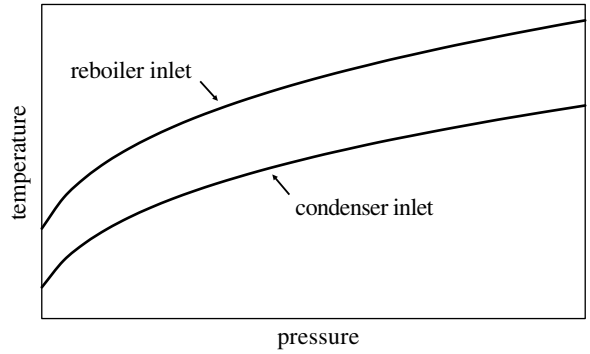

Figure 12.141 *Effect of pressure on reboiler and condenser limits*

hence require the costly installation of a refrigeration system. Similarly designing the column to operate at a higher pressure may reduce the cost of the condenser but increase the cost of the reboiler.

While these decisions are outside the scope of the control engineer, there still remains an opportunity to adjust pressure within the constraints of the installed equipment. Again it is tempting to believe that pressure should be minimised to exploit any condensing capacity and so minimise energy consumption. While this may be true on some columns, it does impose an artificial constraint on feed rate. It may be more economical to raise the pressure to increase the unit capacity. Or recovery of the more valuable product may be increased.

Figure 12.141 also shows the impact on the reboiler inlet temperature. If increasing pressure is economically attractive, then doing so will reduce the temperature difference between the heating medium and the bottoms draw-off. This could result in the reboiler constraining the process. For example if, in our case study LPG splitter, we fix the reboiler at its limit and vary pressure we obtain the result shown in Figure 12.142. With the energy balance scheme in place, increasing the pressure reduces distillate yield and increases the C_3 content of bottoms. However, there is a maximum pressure above which the bottoms product will be off-grade. Since the energy cost is constant, changing pressure simply shifts yield between distillate and bottoms. If the bottoms product is more valuable, then we should maximise the pressure

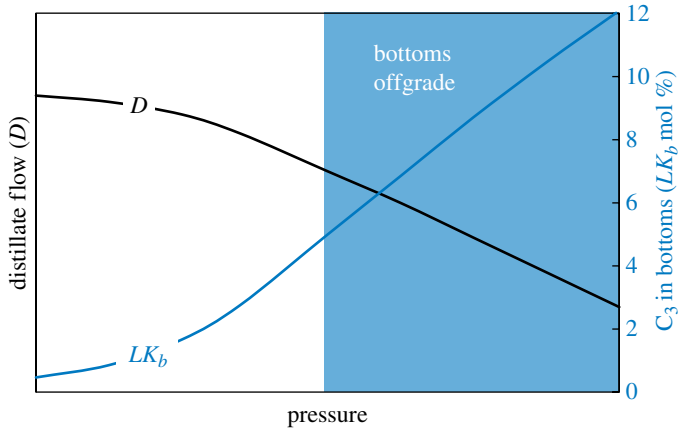

Figure 12.142 *Effect of pressure on product yield*

within the composition target. If the distillate product is more valuable, then we should minimise pressure until some other constraint is reached, such as condenser duty.

The economic value of optimising pressure can vary greatly. On our LPG splitter, exploiting the improvement in relative volatility to reduce energy consumption (as in Figure 12.140) will bring relatively little profit. However, if there is a significant difference in product prices, then exploiting it to improve yield (as in Figure 12.142) can be very lucrative. Of course, on large energy users such as our C3 splitter, even a small percentage energy saving is very attractive.

Should the economics dictate that pressure should be minimised, in order to automate this, some measure is required of how close the current pressure is to minimum. This is most commonly detected by monitoring the output of the pressure controller. For example, if control is achieved by manipulating coolant flow, the controller output might be maintained at 90% by automatically adjusting the pressure SP. Assuming a well-calibrated valve then this ensures the valve will be about 90% open thus leaving sufficient spare condenser capacity to control pressure during minor disturbances. The same approach can be taken whatever the pressure controller manipulates but the use of other indicators can give earlier warning that the minimum pressure has been reached and so permit closer operation to the constraint. For example, if there is a flooded condenser, controlling against a minimum liquid level in the condenser might be more effective. If the pressure controller manipulates coolant flow, under circumstances where the increase in coolant temperature is small, Smith [35] suggests monitoring the ratio between the increase in coolant temperature across the condenser versus the difference in temperature between the overhead vapour and the coolant supply. Keeping this ratio above a value of around 0.2 should ensure that the condenser limit is not reached.

12.22 Energy/Yield Optimisation

In the same way that pressure can be manipulated to exploit the difference in product prices, so can target compositions. We saw in Section 12.1 how cut and separation can be adjusted to maximise revenue. The feasible operating region developed for these variables can be converted to one based on the actual MVs. For our case study LPG splitter, with the energy balance scheme in place, the MVs available for composition control are reboil and reflux. Figure 12.143 shows the feasible operating

region. Note that this is not exactly the same as that applicable to MPC (Figure 12.129) which assumes linear behaviour.

If both products are of equal value, there is no point in producing one at the expense of the other. The object should be to minimise energy consumption by making both products exactly at their purity specifications, i.e. operate where the constraints in Figure 12.143 intersect.

If, however, propane is more valuable than butane we still want to produce propane exactly on specification (to maximise the C_4 sold at propane prices) but may wish to operate with giveaway against the butane specification. If the additional C_3 recovered from bottoms is more valuable than the additional energy cost we want to move away for the intersection of the constraints along the line of constant HK_d. The decision we have to make is how far we move along this line. We covered this as part of the MPC case study described earlier in this chapter. However, if the economics dictate, MPC will move along this line until another constraint is reached. It cannot detect that it may have passed through a point of maximum profitability.

Figure 12.144 shows the effect of adjusting reboiler duty. A composition controller on propane, adjusting reflux, keeps the distillate composition constant but there is no control of bottoms composition.

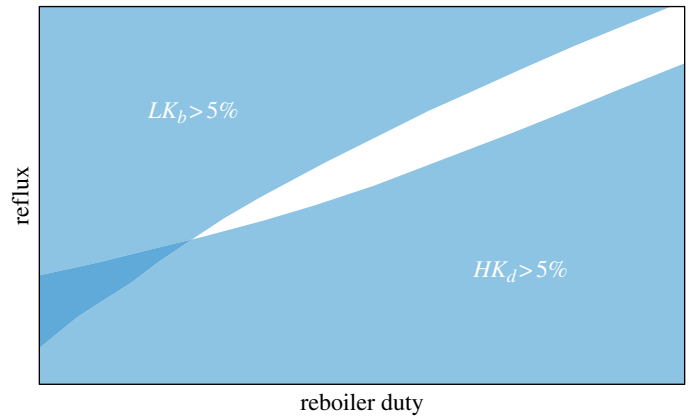

Figure 12.143 *Feasible operating region*

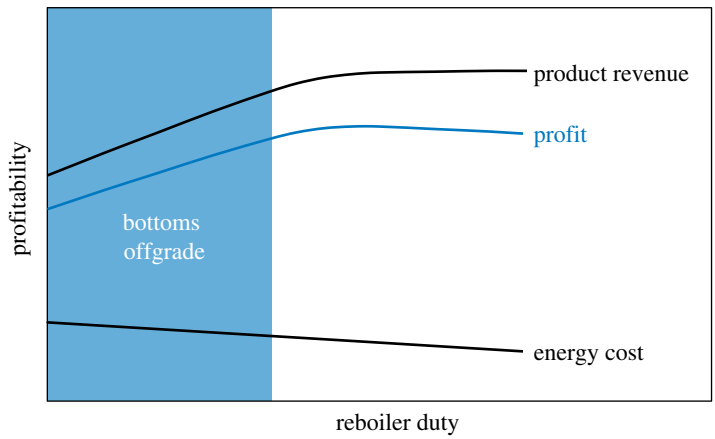

Figure 12.144 *Effect of reboiler duty on profitability*

As duty is increased the energy cost increases linearly – resulting in a decrease in profitability. However, the line of product revenue initially rises but soon flattens. MPC, based on linear correlations, cannot take account of this. As the bottoms purity increases there is little C_3 left to recover and so increasing reboiler duty has little effect on yields. There will be a point at which the additional propane yield does not justify the use of additional energy. This is the point of maximum profitability. Taking the energy cost from the product revenue gives the profit curve, which reaches a maximum at this point.

This point of maximum profit exists in theory on all distillation columns. However, it may not be an attainable point. For example, as shown in Figure 12.144, reducing reboiler duty will eventually cause the bottoms product to go off-grade. In our case the maximum is not beyond this constraint, but that will not be the case for all columns. Indeed if we were to reduce the price difference between propane and butane, the peak would move into this off-grade region. We may also reach other constraints that place an upper limit on reboiler duty – such as a limit on the reboiler itself, the condenser or the column. Under these circumstances the most profitable operation would be at this constraint.

Figure 12.145 plots the profitability curve plotted against the bottoms composition. It illustrates how much giveaway there is, against the specification of 5% C_3, when at optimum operation.

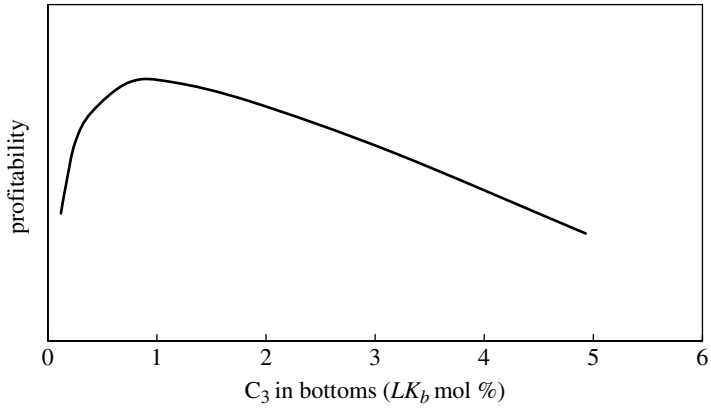

Figure 12.145 *Effect of butane giveaway on profitability*

13

APC Project Execution

13.1 Benefits Study

The cost of process control, as a fraction of the total construction cost of the process, has risen substantially since the early 1960s. Then it was around 5%, now it is closer to 25%. In the 1960s the view was that some instrumentation was necessary but costs should be kept low. As a result plants had the minimum of measurements – just enough for safety and operability. Much of the instrumentation was local to the process, not repeated in the control room, and most of the controllers were single loop with the occasional cascade controller.

Processes are now much more extensively instrumented, with most of it in the control room. The instrumentation has become *smart* supporting a wide range of features such as linearisation, alarms, self-diagnostics and networking. The control buildings have become far more sophisticated with blast-proofing, climate control, ergonomic design and specialist lighting. The control systems have progressed from local and panel mounted controllers to DCS with operator consoles and links to supervisory control computers and data collection systems. Data from other sources such as the laboratory, product storage and scheduling department are increasingly integrated with the control system. Sophisticated control applications based in the DCS and using MPC packages are now commonplace. Rigorous equation based closed loop real-time optimisers are installed on a wide range of processes.

While some of the increased investment was driven by higher safety standards, increased environmental concerns and greater awareness of the value of process data, a large proportion of the justification derives from improved process control. In this section we focus on how those benefits can be quantified.

The management of some manufacturing companies are so convinced of the value of improved control that they require only the most cursory examination of the benefits. While control engineers clearly welcome this approach, it is not without risk. A detailed study would ensure that reality matches expectation; if it does not then it is far better to disappoint before large costs are incurred. It also supports decisions as to whether more costly options should be included. For example does the incremental benefit, of the use of an on-stream analyser versus an inferential, justify the incremental cost?

An unfortunate fact of life is that managers move jobs. By the time it is underway the manager who fully supported the project as a 'no-brainer' may be replaced by one more sceptical. While the decision to progress the project is unlikely to be reversed, there is a danger that the necessary ongoing costly support will be allowed to wane. The study is an opportunity for everyone involved to 'buy in' to the benefits; this is important for the long term success of the project.

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

A reasonable level of accuracy in benefit estimate should be sought. Calculations should be based on actual process performance rather than global statements such as '2% increase in capacity' or '1% energy saving'. Underestimating benefits could result in an attractive project not being sanctioned. Overestimating them will get the project approved but if the benefits later captured fall well short of what was claimed, it is unlikely that the appropriation request for next project will be given much credibility. The best tactic is to slightly understate the true value – both as a contingency against possible implementation problems and to generate kudos when the claimed benefits are exceeded.

It can be tempting to delay benefit studies if process revamps are being explored or in progress. In most sites this is often a permanent state and so the study could be delayed by many years and miss opportunities to capture benefits. Experience shows that the total benefits are reasonably robust to changes not only in the process, but also in the site economics. The operating constraints may change but the process operator's ability to reach each constraint remains unchanged. The benefits will still be captured but perhaps not in the way initially envisaged. The advantage of MPC is that it can readily be reconfigured to deal with such changes in circumstances.

It is important to select the right team to execute the study. Resourcing it entirely in-house carries risk. As described in Chapters 8 and 12, it is common for controllers to be configured to better achieve the current operating strategy – even if that strategy is incorrect. An outsider, properly examining the process economics, is more likely to challenge existing operating strategies. If the argument to change is convincing there is no need to wait for the project. Early manual implementation is likely to capture benefits larger than those attributable to improved control. An outside specialist is also more likely not only to know what newer technologies are available but also, more importantly, how successful they have been elsewhere.

One might consider getting the MPC implementer to conduct the benefits study. The argument often presented for this approach is that whoever calculates the benefits should be able to guarantee them if they are also responsible for their capture. Indeed, this is often the argument presented by implementers, suggesting they might not be held accountable to capture benefits quantified by others. The statistical techniques used to determine the benefits are common to all organisations. It is extremely unlikely that an implementer would decline a project if a requirement was to deliver the process improvements quantified by another using the same methods as his own.

Further, the single-supplier approach is likely to force the owner to compromise on the best choice of provider. It is often not the case that the best implementer is also the best at performing benefits studies. Given the relative size of the pieces of work, the owner will choose in favour of implementation skills and risk overlooking potentially valuable profit improvement opportunities.

The risk of having the implementer conduct the benefit study is that there is no guarantee that it will be executed impartially. Deliberately steering the project definition to favour the implementer rather than the owner is not in the implementer's long term interest; it will eventually harm their reputation. But the implementer will look for benefits that can be captured by the technology he offers, rather than those which might require technology from a competitor or with which he is unfamiliar. While some implementers will claim to be 'independent', in that they are not tied to a particular technology supplier, they still have a very strong vested interest in maximising the profit they will make from implementation.

13.2 Benefit Estimation for Improved Regulatory Control

At the regulatory control level there are the obvious benefits of keeping the process stable, safe and operable by a reasonable number of process operators. The process design team will most likely have already met these requirements, so we take this as our base case above which we have to economically justify further costs.

Addressing first the basic regulatory control layer, we should recognise that if higher levels of control are economically justified then, even if the basic controls capture no benefit in their own right, they provide the necessary foundation for these higher levels. Indeed, it was commonplace for APC to be included in a re-instrumentation project to provide the justification for the improved basic control. The problem is that this can create the culture that basic control is a 'necessary evil' and that its cost should be minimised. This is in conflict with a culture of maximum profit. There remain many sites where basic instrumentation is neglected, despite large investments in APC. Good basic control should not only be recognised as necessary for APC, but also that it can be valuable standalone.

Good regulatory control permits faster changes of operating conditions. Processes will often have different operating modes. They may process different feeds or produce different grades of product. On many processes the change of mode is the only significant disturbance. Mode changes can be costly. When switching between product grades, the material produced before the new specification is reached will have a lower value. It may require storage, reprocessing or is downgraded to a lower value product. If we can shorten the mode change then less of this material is produced. Further, because we produce more of the target product in the same time period, we have effectively debottlenecked the process. This is particularly true if off-grade material has to be reprocessed – using up valuable capacity and increasing operating costs. We can reduce storage costs by moving closer to a make-to-order process. And we may be able to gain market share by being a more reliable and flexible supplier.

Improved regulatory control can reduce maintenance costs. It is self-evident, if pumps and compressors are subject to fewer pressure surges, temperature deviations, etc., then bearings, seals and gearboxes are less likely to fail. While it is virtually impossible to predict these savings there have been cases where the *mean time between failures (MTBF)* for machines in difficult services have increased by a factor of three. If the equipment is critical, in that its failure requires a process shutdown or turndown, then not only are maintenance costs reduced but process capacity is increased.

Improved control makes better use of the process operator. Gone are the days where process control projects were justified in terms of a reduction in manpower. Most processes now have manning levels close to the minimum required for routine start-ups and shutdowns, and to deal with emergencies. Nowadays improved control makes the operator's role less mundane. Instead of being fully occupied keeping basic variables at desired conditions, more time can be spent on improving the operation. This of course has an impact on the quality of the personnel required and their level of training, but a good operator is wasted if employed in a mundane role.

A common way of estimating benefits for regulatory control is to quantify how much closer a process constraint may be approached. Figure 13.1 shows the classic drawing. With poor control the process operator will enter a controller SP such that worst deviation does not violate the constraint. Improving the regulatory control and reducing the deviations increase operator confidence that the controller can better handle process upsets and so the SP can be moved closer to the constraint.

This type of analysis should only be applied to a hard constraint – one which should only be approached from one side. An example is maximising pressure against a trip or relief valve setting. It should not be universally applied to assessing the benefit for more closely approaching, for example, the specification of a liquid product. If the product is routed to storage before being sold or processed further then its specification is a *soft constraint*. It can be temporarily violated and, provided corrective action is taken and the product tank well-mixed, the finished product can still be produced exactly on grade, no matter how large the deviations from SP. Indeed, a frequent oversight in benefit estimation is to compare the limiting property of a product being sold against its specification. It is tempting to believe that if the product is exactly on specification then there is no benefit to be had. However, if the product has been produced on a process not well controlled so that the rundown property is highly variable, large savings may still be possible.

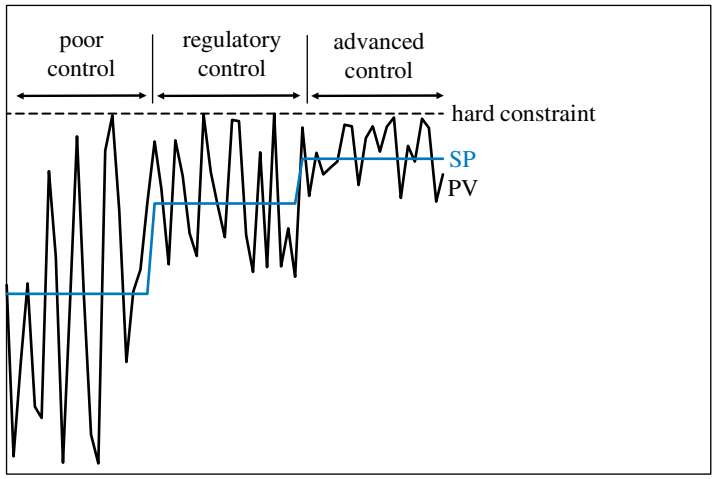

Figure 13.1 *Benefit of improved control versus a hard constraint*

Most processes are fundamentally nonlinear. If we examine some of the key equations governing process behaviour, this quickly becomes apparent. Heat transfer is fundamental to almost every process. Whichever way this is achieved involves nonlinearity. For example, in Chapter 10 we applied Stefan's Law to estimating benefits on a fired heater. The law states that the rate of heat transfer (Q), in the radiant section of a fired heater, varies highly nonlinearly with the temperature (T).

$$Q = \sigma T^4 \tag{13.1}$$

In the convection section the *Five Fourths Power Law* applies

$$Q \propto T^{5/4} \tag{13.2}$$

While linear with respect to heat transfer coefficient (U) and area (A), the rate of heat transfer in a heat exchanger is highly nonlinear with respect to temperature. This is governed by the log mean temperature difference (LMTD). So, for example, if the flow though one side of a heat exchanger is doubled we do not achieve double the heat transfer.

$$Q = \frac{UA\left(\Delta T_{in} - \Delta T_{out}\right)}{\ln\left(\Delta T_{in} / \Delta T_{out}\right)} \tag{13.3}$$

Chemical reactions are governed by Arrhenius's Law which states that the rate of reaction (k) is proportional to an exponential function of absolute temperature.

$$k \propto Ae^{-E/RT} \tag{13.4}$$

Key to the distillation process, Antoine's Law relates vapour pressure (P_0) to an exponential function of temperature.

$$\ln\left(P_0\right) = A - \frac{B}{T + C} \tag{13.5}$$

Even mixing is not immune to nonlinearity. If blending on a weight basis, SG does not blend linearly. Blending two streams of flows F_1 and F_2, with specific gravities SG_1 and SG_2 gives a combined SG of

$$SG = \frac{(F_1 + F_2)SG_1SG_2}{F_1SG_2 + F_2SG_1} \tag{13.6}$$

Combining streams of different viscosities is governed by the *Refutas Equation* which converts viscosity (v) to a linearised *viscosity blending number (VBN)*.

$$VBN = 14.534.\ln\left(\ln\left(v + 0.8\right)\right) + 10.975 \tag{13.7}$$

The VBN is determined for each stream, blended on a weight basis, and the result converted back to viscosity.

Figure 13.2 shows the upshot of all the sources of nonlinearity. As almost any process parameter is varied the effect that it has on operating cost is nonlinear. Thus the cost of manufacturing a product always exactly on specification is less than that of producing it when only on average it is at specification.

Figure 13.3 shows an example of the calculation that might be performed on a very high purity LPG splitter. The distillate specification is 0.05 mol% C_4 in propane. The material sold from storage has a concentration very close to this specification and so there would appear little benefit to be obtained from better controlling it. However, the rundown composition of the stream to storage varies between 0.02 and 0.08%. Figure 13.3 shows the reboiler steam consumption per unit of propane product. This relationship could be obtained from historically collected process data or from process simulation. It shows that steam usage varies from 1.10 to 1.74, giving an average of 1.42 t/t propane. By halving the variation in C_4 content to the range 0.035 to 0.065%, the steam variation reduces to 1.16 to 1.42, giving an average of 1.29 t/t propane. The saving of 0.13 t/t propane, on a typical size column, would easily justify the cost of an on-stream analyser and the control application.

With the advent of MPC the split of benefits between regulatory control and constraint control became less clear. Traditional APC applications, such as those controlling product composition, became part of MPC. But both exploit the 'comfort zone' left by process operators. They achieve this by taking action far

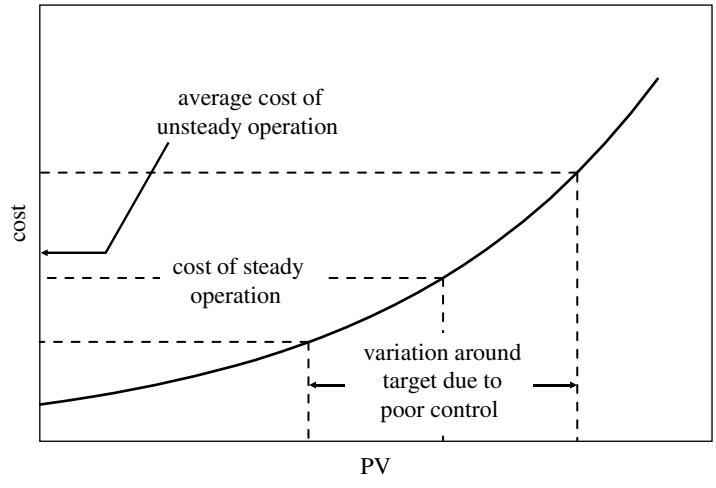

Figure 13.2 *Benefit of improved control versus a soft constraint*

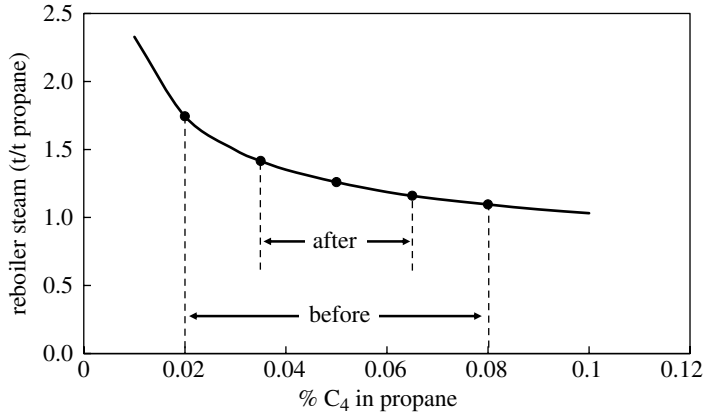

Figure 13.3 *Benefit of improved control of product purity*

more frequently than even the most attentive process operator could achieve. So they will rapidly exploit even a transient opportunity to more closely approach a target. And they will avoid costly violation of constraints. Constraint control, as it is now defined, captures by far the largest portion of the benefits available on continuous processes. On batch processes, constraint control usually has only limited application; the larger benefits are more likely to be attributable to scheduling techniques.

While MPC will capture additional benefits associated with its choosing the optimum constraints to approach, like its predecessors it will approach these more closely by reducing the variation in the deviation from target. The assumption that the variability or standard deviation (σ) is halved by the implementation of improved regulatory control has become a de facto standard in the process industry. It has no theoretical background; indeed, it would difficult to develop a value theoretically that is any more credible. Post-implementation audits usually confirm that this assumption is realistic.

Thought should be given to the data collection interval. Laboratory results are typically collected at an interval of 8, 12 or 24 hours. This will generally be much longer than the time the process take to reach steady state and it is realistic to assume that a controller, typically executing every minute, will improve control of quality. However, the same controller would probably not be fast enough to cope with a process with very short dynamics – such as compressor surge protection. This would require a much shorter execution frequency and so the data collected to estimate the potential process improvement should similarly be collected at a much higher frequency – with an interval of perhaps a few seconds.

We also need to consider the number of data points collected. This should be no less than 30 values for there to be any confidence in the estimate of standard deviation. However, in the case of laboratory samples taken several times a day, this might only cover a week or so. It would be advisable to analyse around three months of laboratory data, which may involve around 300 results. With modern data collection systems and spreadsheets, this would entail little additional work.

Laboratory data is often available both for the product *rundown* (the product leaving the unit) and for storage (the product exported). Which data should be used depends on the type of benefit being analysed. For example, assessing the benefit of improving the control of a highly nonlinear process (such as that illustrated in Figure 13.3) would require the use of rundown data. The benefit of reducing the giveaway in a product sold to a customer should be based on tankage results – often certificates of quality.

Some judgement should be exercised in assessing which data points are suspect and therefore classed as *outliers*. If they are genuine deviations, that might reasonably be dealt with by the proposed control scheme,

their inclusion will increase the estimate of benefits that would be captured. However, the inclusion of values arising from major process disturbances such as those caused by equipment failure, or the inclusion of data that are incorrect, will result in an overestimate of the worth of the scheme.

While process measurements are generally normally distributed it is wise to check that this is the case before calculating the standard deviation. Figure 13.4 illustrates a situation where the distribution is clearly not normal. Known as a *bimodal* (because it has two *modes*) distribution, the two peaks occur in this case because the unit has two operating modes – one where the target is 100 and another where it is 120. In both cases the standard deviation is 5. Blindly calculating the standard deviation for the combined data would give a result of 11 – more than doubling the benefits that would be claimed. In effect the distribution is assumed to follow the dashed line rather than the solid one. Most spreadsheet packages include statistical functions that can be used to determine how close a distribution is to normal. These include *skewness* and *kurtosis*, as described in Chapter 14. Acceptable ranges for each of these is −0.5 to +0.5. In this example the skewness is 0.28 but kurtosis is −1.18. This indicates that the distribution is much flatter than we would expect and so should prompt further investigation. Alternatively the *quantile plot* (Figure 13.5) technique, as described in Chapter 14, can be used. The cyclic deviation, from what would be expected if the distribution were normal, indicates that there are probably two peaks in the distribution.

In our first example the two operating modes are far apart. Consider now the distribution shown in Figure 13.6 where the two targets are 100 and 110. It would be tempting to treat the distribution as *unimodal* but doing do would result in overestimating the benefits by 40%. The skewness is 0.083 and kurtosis −0.54 – only just outside the acceptable range. The quantile plot (Figure 13.7) also barely shows the problem, illustrating the need to review carefully the meaning of the data collected.

If there are multiple operating modes then one approach is to analyse each separately by ensuring only data collected during each mode are included. Alternatively, if the average standard deviation for all modes is required, all the results should first be converted to deviations from target.

Consideration should be given to the impact that measurement error might have. While generally not a major issue with most instrumentation it can be a problem with laboratory data. Some laboratory tests are subject to inaccuracies that can be of the same order of magnitude as the true variation in the product quality. Since improved control can do nothing to improve the laboratory, even perfect control may not halve the reported standard deviation. Most laboratory tests have a quoted repeatability (r).

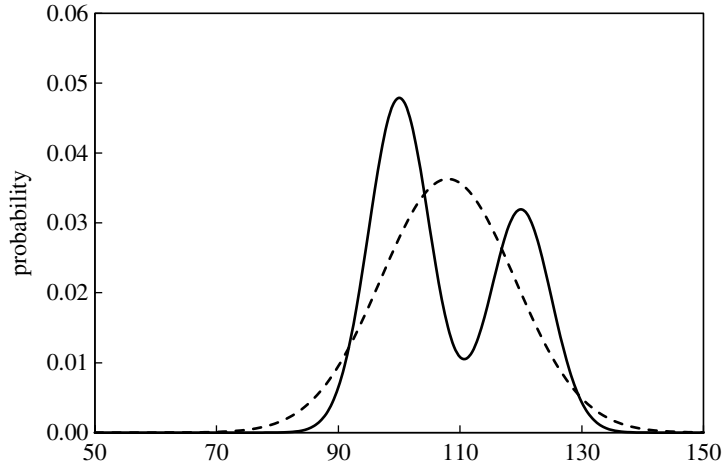

Figure 13.4 *Effect of two well-spaced operating modes*

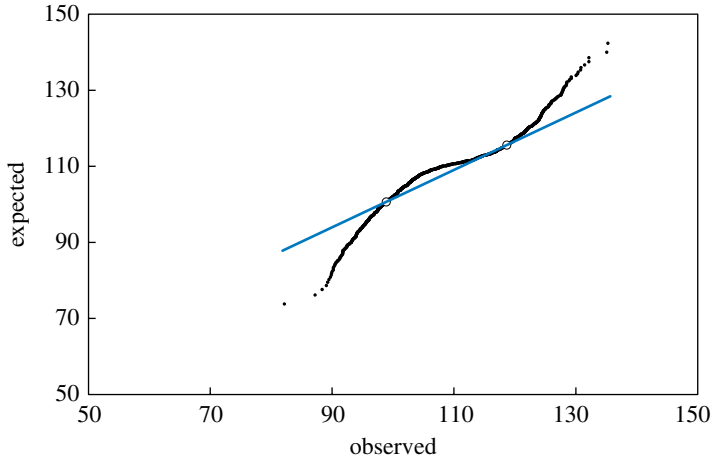

Figure 13.5 *Quantile plot for well-spaced operating modes*

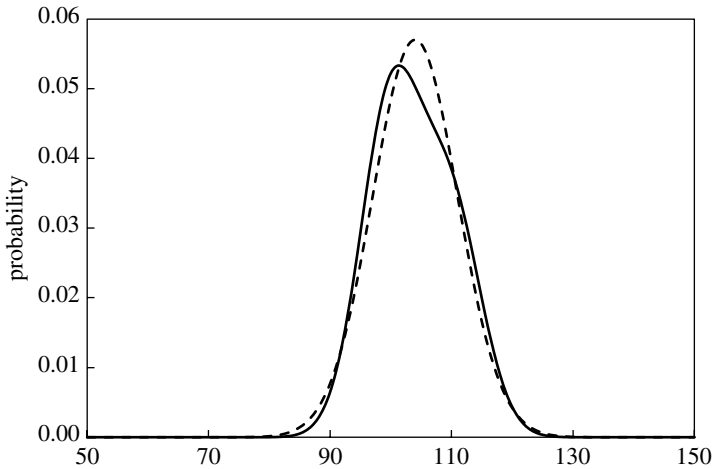

Figure 13.6 *Effect of two closely spaced operating modes*

This is usually defined as a multiple of the standard deviation (σ) of the test result. For example, the ASTM defines it as 2.8σ – as explained in Chapter 14. The standard deviation of the 'true' quality could then be determined.

$$\sigma_{true} = \sqrt{\sigma^2_{reported} - \left(\frac{r}{2.8}\right)^2} \tag{13.8}$$

However, if adopted, such an approach should be used with caution. Laboratory repeatability is often better than the quoted standard. Or the inaccuracy may be increased by other sources of random error – such as poor sampling procedures.

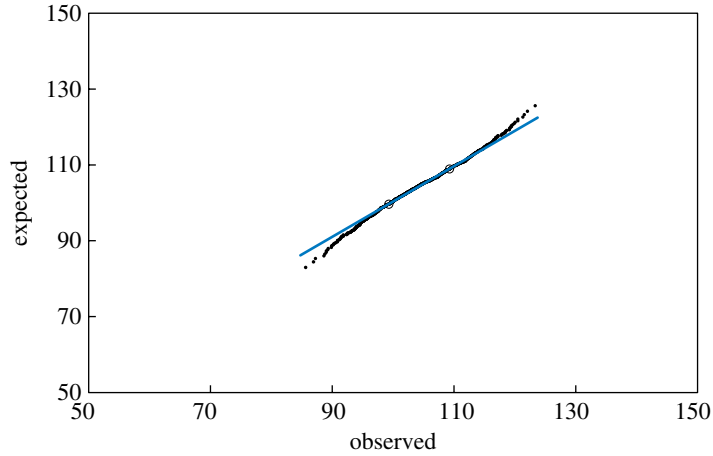

Figure 13.7 *Quantile plot for closely spaced operating modes*

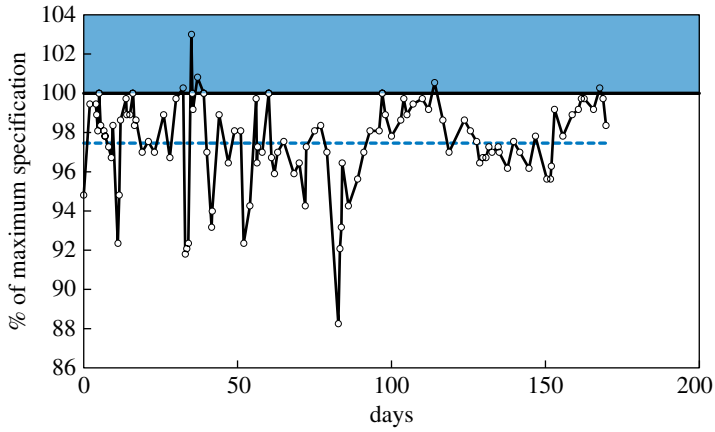

Figure 13.8 *Process performance against a maximum specification*

There are a variety of ways in which the benefit of reducing the standard deviation can be assessed. The *Same Percentage Rule* [36, 37] is based on the principle that if a certain percentage of results already violate a specification, then after improving the regulatory control, it is acceptable that the percentage violation is the same. Figure 13.8 shows six months of data collected for a benefits study. The results have been normalised to percentage of specification. There is an economic incentive to approach the limit of 100% as close as possible. The chart shows that, in doing so, several results (in the shaded area) are off-grade. The mean value is 97.5% (shown as the dashed line) and the standard deviation 2.3%.

Skewness is −1.1, indicating that negative deviations from the mean tend to be greater than positive deviations. This could be explained by the operator having given, not surprisingly, more attention to off-grade production than giveaway. Kurtosis is 2.1, indicating that the distribution has long tails compared to the normal distribution. Figure 13.9 shows the quantile plot for the same data. While one point would

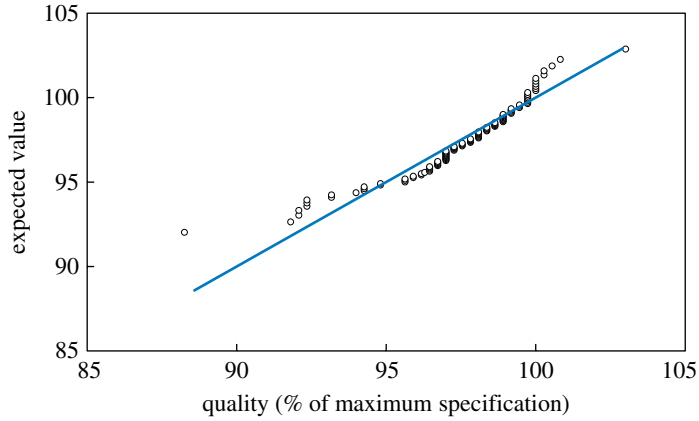

Figure 13.9 *Quantile plot for process data*

appear to be an outlier and could perhaps be ignored, it confirms that the distribution is only approximately normal – especially in comparison to the example, known not to be normal, described by Figure 13.7. However, these validation techniques should be used only to prompt more detailed investigation as to why the distribution is not exactly normal. If there is no reason then, given all the other approximations made, treating the distribution as normal should cause no problem.

Figure 13.10 shows the same data as a cumulative frequency plot. The solid black line is the best fit normal distribution. If improved control halves the standard deviation (coloured line) then the mean can be increased by half the current mean deviation from target, i.e.

$$\Delta\bar{x} = 0.5\left(x_{\text{target}} - \bar{x}\right) = 0.5\left(100 - 97.5\right) = 1.25 \tag{13.9}$$

In our example, improved control will increase the mean result from 97.5 to 98.75. Figure 13.10 shows that this gives the same percentage violation of the specification.

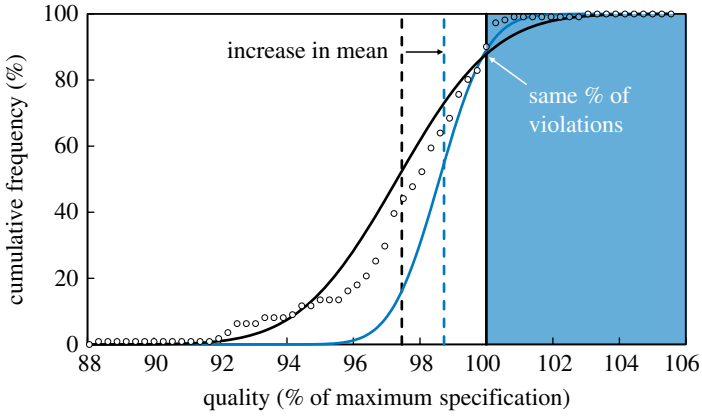

Figure 13.10 *Same percentage rule*

The *Same Limit Rule* [36, 37] is used when a large number of results violate the specification. In this case we choose a new target which is violated by a reasonable proportion of results. The resulting change in the mean is given by

$$\Delta\bar{x} = \frac{\sigma}{2} P(m) \tag{13.10}$$

The value for *P(m)* is taken from Figure 13.11. For the normal 5% violation the value is 1.64; for the more demanding 1% it is 2.33.

Figure 13.12 shows data for a different product quality from the same study. This time the specification is a minimum and again there is an economic incentive to approach the limit of 100%. In this case around half the results lie in the off-grade shaded area. Indeed, the mean is below the limit at 97.9%. The standard deviation is 9.1%. Figure 13.13 shows the same data plotted as a cumulative frequency plot. The solid black line is the best fit normal distribution. Having chosen a value for *P(m)* to give 5% violation of the new minimum limit, the mean is reduced to 90.4%.

Once the process improvement is quantified in engineering units, it has to be converted to a profit increase. Care has to be taken in identifying the effect of such marginal changes. For example, on our case study LPG splitter, increasing propane yield by recovering more C_3 from the butane product might appear to involve quite simple economics. This is the case if the additional propane is sold to a customer and another customer receives less butane. We need only to consider the selling price of both streams and perhaps any change in reboiler energy consumption. But it may be that we have some internal use for butane. For example butane is often a blend component in gasoline. If less is available then we may sell less gasoline so it might be tempting to use the gasoline selling price in our calculation of profit. However, we also have to consider the change that we have made to the butane composition. By reducing its C_3 content we have reduced its RVP (Reid vapour pressure), which then reduces the RVP of the blended gasoline. Since gasoline is often blended to a maximum RVP specification we might choose to inject more butane into gasoline and, as a result, sell more gasoline and even less butane to our customer. The situation is complicated by the fact that butane is traded on a weight basis and gasoline on a volume basis. Our calculations therefore need to take into account the change in the density not only of the additional gasoline sold, but of all the gasoline sold.

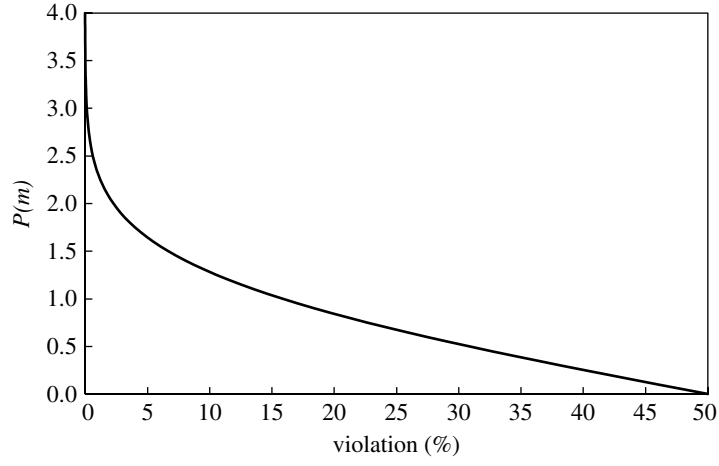

Figure 13.11 *Cumulative probability function of the standard normal distribution*

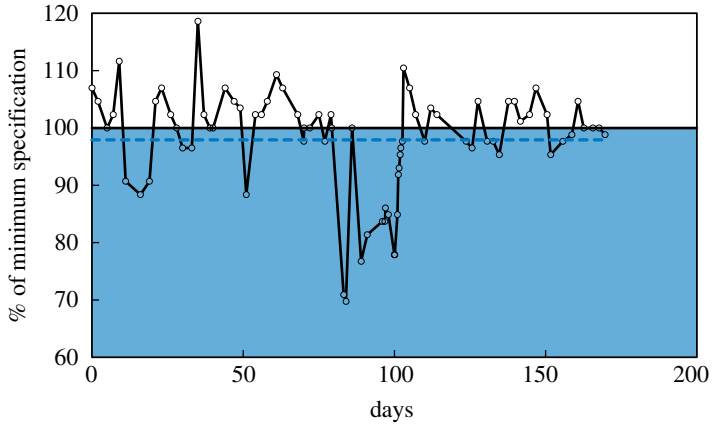

Figure 13.12 Process performance against a minimum specification

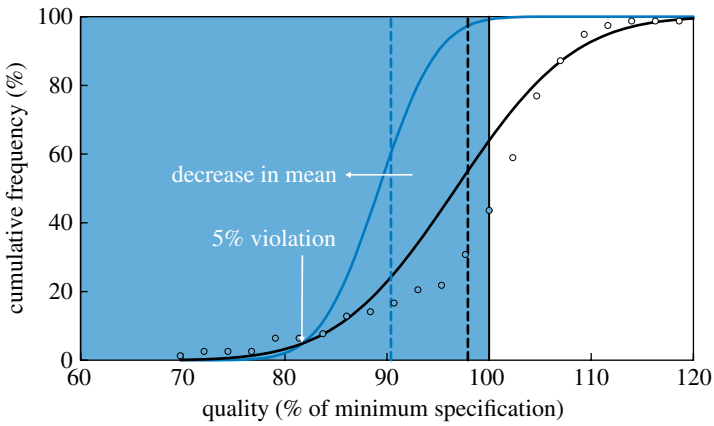

Figure 13.13 Same limit rule

Describing all but an example of the complexity of process economics is beyond the scope of this book. The example makes the point that the engineer should consider the full impact of the change being made. Obtaining marginal prices from the planning and economics section is usually too simplistic. A more rigorous approach will occasionally reveal that an assumed operating objective is incorrect and that meeting such an objective more closely with improved control will lose money.

13.3 Benefits of Closed-Loop Real-Time Optimisation

Estimating benefits for closed-loop real-time optimisation (CLRTO) requires particular attention. We have to distinguish CLRTO from what can be achieved with MPC. MPC is based on simple process models developed empirically. They support simple process economics and, if properly configured, will locate a constrained optimum. CLRTO is much more rigorous and is based on first-principle nonlinear engineering

models. Its implementation can be extremely costly. The process simulation is often equation-based. For complex processes it is not unusual for many thousands of equations to be configured. While many come pre-packaged from a library of equipment types they have to be checked and calibrated against real process data. This incurs large technology and engineering fees.

In addition to the process simulation and optimiser, other modules are required. The optimiser is likely to be steady-state and therefore should only be permitted to execute if the process is at steady state. Further, if the optimiser takes any significant time to identify optimum conditions, the steady-state detection must also check that the process conditions have not changed since the start of the optimiser cycle. Tuning steady-state detection can be demanding; processes rarely attain exact steady state so sufficient leeway has to be given so that the optimiser runs reasonably frequently but without jeopardising the accuracy of the result.

Data reconciliation is required to handle suspect or missing measurements; if mass and energy balances do not close then the optimiser may fail.

Costly expertise is required to maintain the system, requiring an engineer strong in the technology itself and very experienced with the process and its economics.

There is also the potential for the optimiser to generate a non-optimum solution. An economic value has to be placed on any stream crossing the process simulation's boundary. Often such streams are feeds or products that do not have a clear purchase or sales price. Valuing intermediate products and energy sources can be extremely complex on integrated sites. Choosing a set of economics consistent with more global optimisation requirements is challenging and prone to error.

Optimum operation is likely to be a function of feed composition. On many processes, susceptible to feed changes, the necessary on-stream analyser technology may not exist. The use of laboratory data introduces a delay which could cause the optimisation to be in error for many hours.

The market for rigorous CLRTO applications peaked some years ago; the initial hype has been largely replaced with some level of scepticism, based on the experience of the users. In many cases the optimiser moved the process to a set of constraints and then did nothing more. While capturing substantial benefits, most of these could have been obtained from a considerably less costly off-line process study. In other cases there is no independent means of determining whether the process has been truly optimised. On many sites the technology has fallen into disuse due to a combination of scepticism about its value and the difficulty in retaining high quality technical support.

Having stated the problems there are, however, a number of very successful applications. The key to achieving this is to first properly consider the technical feasibility, the costs and the benefits. Figure 13.14 shows a simplified example of optimising a single PV. As the PV is adjusted the profit passes through a maximum. This is clearly the first condition for CLRTO to be considered; if the maximum is beyond a constraint then it can be located much more cost effectively by MPC. The second condition is that the optimum is not too 'flat'; the current operation may be far from optimum but the economic impact may be small. Thirdly, the optimum must move. There must be changes in key parameters such as feed price, availability and composition; product price, demand and quality specification; or process changes (such as degradation of catalyst or fouling of heat exchangers). The changes must be frequent and significant. If this is not the case then the majority of the benefits can be captured by offline studies, perhaps using the same simulation tool, to develop sets of recommended operating conditions for each scenario. The benefit of CLRTO comes from the incremental value of tracking a moving optimum compared to controlling at an optimum previously defined offline.

If the benefits are insufficient to justify rigorous CLRTO, or if the plant owner would rather capture less of the benefit and not have to support the complex technology, there are a growing number of less rigorous intermediate solutions. These are effectively extensions to MPC but apply *quadratic programming (QP)* rather than *linear programming (LP)*. The fundamental difference is that objective coefficients are applied to a quadratic function of process variables. This permits an unconstrained optimum to be identified. Like MPC,

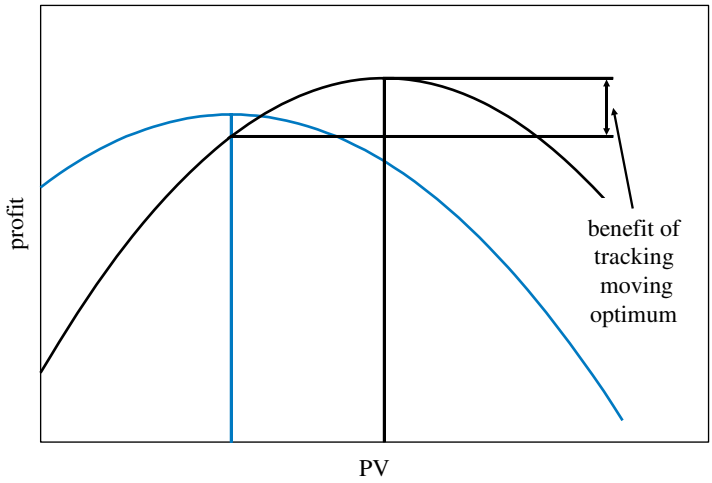

Figure 13.14 *Quantifying benefits for CLRTO*

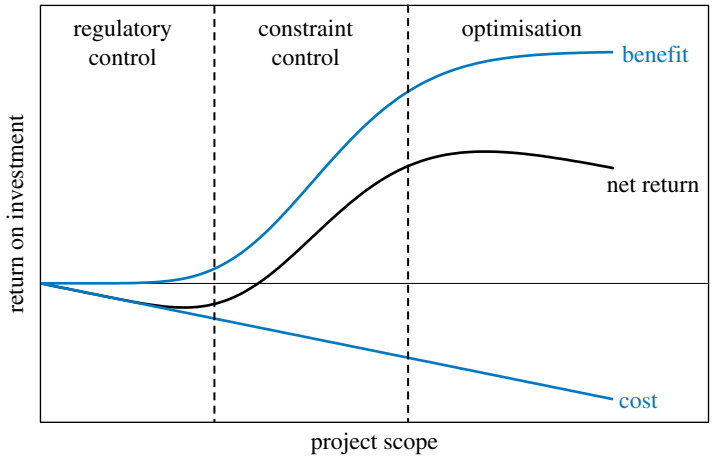

Figure 13.15 *Optimising project scope*

these packages use dynamic models of the process, determined empirically, enhanced as necessary with simple engineering models. Because they do not require steady state they execute far more frequently.

Figure 13.15 shows the effect of varying the project scope. Increasing cost is shown as a negative return. Simple regulatory controls tend to be costly without capturing large benefits in their own right. On existing processes it is common for the instrumentation to need considerable attention, for many of the control algorithms to be changed and tuning constants to more be accurately determined. This can involve significant hardware costs, particularly if it includes upgrading the DCS. It will also involve substantial manpower. It is unlikely that a project of such limited scope would show a strong economic return.

A means of justifying the investment in regulatory control is to enlarge the scope of the project to include constraint control. This is likely to substantially increase the benefits but with usually a more modest impact on cost. The temptation is to progress constraint control without paying much attention to

regulatory control on the basis that, if we can reduce the cost of this layer, then the overall return would improve. But this approach will reduce the benefits captured by the constraint controllers and this lost opportunity will be perpetuated. The cost of re-engineering the constraint controllers to take account of any later improvements to the regulatory controls will probably be prohibitive.

The benefits captured by optimisation are very process-specific. They can match or exceed those captured by constraint control or they can be a very minor addition. The case illustrated in Figure 13.15 is one where a rigorous simulation-based optimiser would probably not be justified but perhaps a much lower cost empirical technique, capturing a portion of the available benefits, might be worthwhile. Since the choice of technology and implementer are likely to be influenced by the scope of the project then it is important to decide upon this first.

13.4 Basic Controls

One of the main aims of this book is draw attention to the improvements possible at the basic control level and to emphasise the importance of completing these before embarking on a MPC project. The choice of control algorithm, its tuning and any additions such as feedforward control or deadtime compensation, affect the overall process dynamics. Once step-testing for MPC has been completed then any change to the basic controls would be very unwelcome and unlikely to be implemented until other circumstances, such as a process revamp, require step-testing to be repeated.

The first priority is to ensure that the instrumentation is operating correctly. Any controllers often out of service should be identified and the problems resolved. All control valves should be checked to ensure that they are correctly sized and the controller not liable to saturate. The type of valve (equal percentage or linear) should be checked. Calibration of valve positioners should be checked and any mechanical defect, such as excessive stiction or hysteresis, be rectified. Any excessively noisy measurements should be dealt with, if possible, by resolving the problem at source – rather than by the use of filtering. Any existing filter not required to prevent excessive actuator movement should be removed. Resolution of some instrument problems can be delayed waiting for delivery of replacement parts or for a plant shut-down. It is therefore wise to identify any problems as soon as possible so that these are not on the critical path of the project.

Once instrumentation is fully functional then control configuration changes should be completed and tuning optimised. If the schedule permits (or if the work is being completed for a new process) then, for the advantages of adopting a standard approach, all controllers should be addressed. If this is not the case then only those controllers having a significant impact should be checked. While it may be possible to substantially improve the time a flow controller takes to return to SP, a saving of a few seconds will not be noticeable on a process where dynamics are measured in minutes. Those controllers that should be reviewed include all level controllers, most temperature controllers (such as those on fired heaters, distillation column trays, etc.) and any other controller where the process dynamics are relatively slow. Any controller identified by the process operator as problematic should also be addressed – whether or not important to the performance of MPC. This will help greatly with operator acceptance of the whole project.

The preferred control algorithms have been described in earlier chapters but in summary they are:

- signal conditioning to help linearise process behaviour
- filtering to remove noise (and the removal of unnecessary filters), using the least squares filter if the lag introduced by the standard exponential filter is excessive
- proportional-on-PV, derivative-on-PV non-interacting PID algorithm tuned for SP changes using the method described in Chapter 3

- averaging level control where appropriate (using gap controller if flow disturbances are very variable), tight level control otherwise
- ratio feedforward on feed rate (particularly if feed rate changes by more than ±20%)

As an alternative other algorithms, such as bias feedforward and deadtime compensation, can be implemented in MPC – depending on which approach is better for operator understanding and what back-up scheme is necessary if MPC is out of service. It is also possible to move averaging level control from the DCS to MPC. This should only be considered if it is desirable to let MPC select which flow to manipulate (i.e. vessel inlet or outlet) depending on where the process is constrained. The DCS controller will still be required as back-up.

If MPC is to be implemented by a contracted specialist company it is unlikely that including, in their scope of work, a detailed check on basic controls will be successful. Such implementers generally require a much lower standard of basic control performance. They will generally check in any case, during the *pre-test* phase of the project, that all the MVs operate on automatic and generally respond properly. They probably will not consider the performance of controllers not destined to be MVs, even though their tuning may still affect the overall process dynamics. Even if asked to be more rigorous the implementer may not have the skills necessary for this work. Further, since they are usually under budget and schedule pressures, it is not in their interest to identify problems that delay implementation of MPC or require a great deal of their attention. The plant owner should take on this work, long before placing the contract with the MPC vendor, if necessary by bringing in outside expertise.

13.5 Basic Control Monitoring

There are a number of proprietary packages that provide a range of performance monitoring techniques for basic controls. While they can offer a valuable insight, they do so at a cost. There are a number of simple techniques that the engineer should consider first. For example, simply monitoring the time that the controllers are in automatic mode will quickly identify any that have a recurring problem. This might be enhanced by also recording the number of times the mode is changed. It can also be useful to record the number of times controller tuning is modified. Recording the period during which the PV is in alarm and also the maximum deviation from SP might also indicate a control problem. Monitoring whether the PV is 'bad' or at the limit of its range will help identify those instrument ranges that should be increased. Similarly monitoring the minimum and maximum of the actual operating range can identify those situations where a reduction in range might be beneficial. Monitoring the output of the controller will identify whether it is saturated for significant periods, suggesting an increase in valve size might be appropriate. Recording the minimum and maximum output would identify whether a reduction in valve trim would be beneficial.

Criteria used to tune a controller can also be used to assess the performance of an existing controller. This might include, for example, calculating the variance of the error over a defined period. This has somewhat limited value because it is affected by the size of disturbances. The *Harris Index* [38] addresses this by comparing the variance of the error against the minimum achievable. This minimum would be achieved by applying an optimally tuned PID controller with deadtime compensation (the Smith predictor as described in Chapter 7).

$$HI = \frac{\sigma^2_{error}}{\sigma^2_{minimum}} \tag{13.11}$$

This index will vary between 1 (for the best achievable) and very large values (for poor control). Perhaps more intuitive is to use the reciprocal in that it varies from 0 (for very poor control) to 1 (for the best achievable). While these indices are immune to the size of disturbance, most controllers do not have

deadtime compensation and so minimum variance becomes a theoretical concept rather than what can be achieved in practice. Others have therefore suggested techniques based on what might be achieved by, for example, an optimally tuned conventional PID controller.

The penalty functions defined in Chapter 3 can also be adapted for monitoring purposes. These too are affected by the size of disturbance and so, for example, *IAE* is normalised by dividing by ΔSP – taking care to use the same unit of measure as that used for error (E) in calculating *IAE*. The result will then have units of time; to make it dimensionless it is also divided by the lag (τ) – again taking care to use the same unit of time as the controller scan interval that was used to calculate the penalty function.

$$IAE' = \frac{IAE}{\Delta SP.\tau} \tag{13.12}$$

The dimensionless ISE is given by:

$$ISE' = \frac{ISE}{\left(\Delta SP\right)^2.\tau} \tag{13.13}$$

As we have seen in Chapter 3, the preferred penalty function is *ITAE*. Before considering how this might be used for controller monitoring, because settling time (t) is minimised by minimising *ITAE*, it can instead be used to assess performance. Rearranging Equation (3.124)

$$\frac{t}{\tau} = 2\frac{\theta}{\tau}\left(\frac{\theta/\tau + 2}{\theta/\tau + 1}\right) \tag{13.14}$$

Typical values of these indices, when minimum variance has been achieved, are given in Figure 13.16. To use *ITAE* as an ongoing measure of controller performance a more rigorous approach is justified. As with *IAE* and *ISE*, we need to normalise *ITAE* by dividing by ΔSP. Figure 13.17 shows how the minimum achievable $ITAE/\Delta SP$ varies with process dynamics. It is drawn for the preferred I-PD algorithm assuming a 15% limit on MV overshoot. As we saw in Chapter 3, removing this constraint has no effect if the

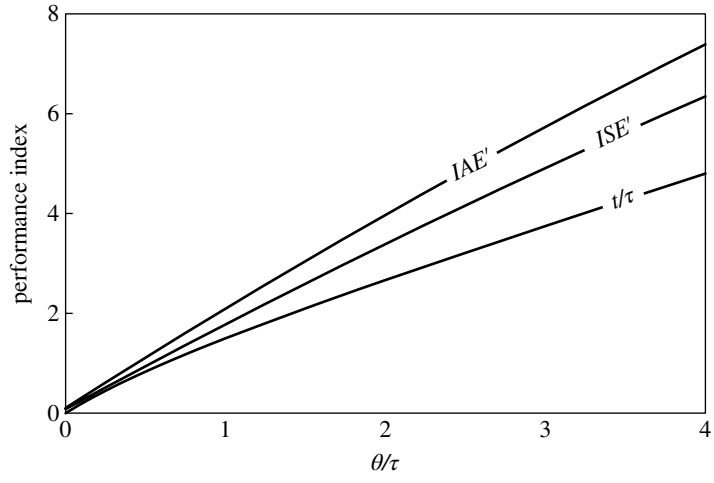

Figure 13.16 *Performance indices for minimum variance*

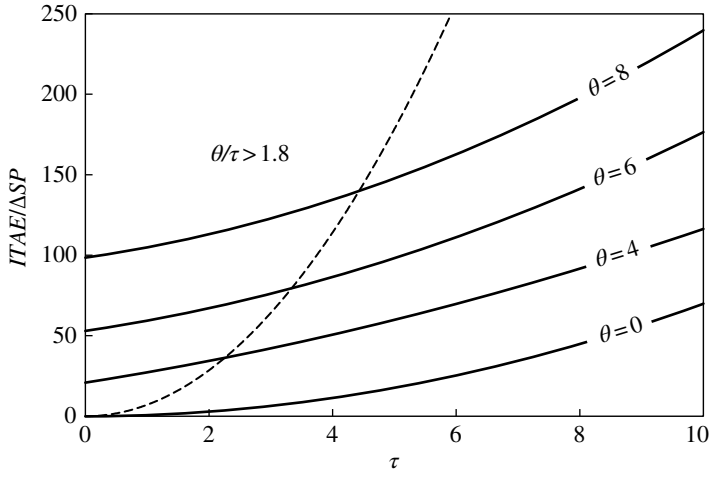

Figure 13.17 *Impact of dynamics on minimum achievable ITAE*

θ/τ ratio is greater than 1.8 (the region shown to the left of the dashed line). As shown by Equation (3.122) the calculation of *ITAE* involves both the controller scan interval and the elapsed time since the change in SP. If the process dynamics are measured in minutes then the parameters used to calculate *ITAE* should also be in minutes. The parameter *ITAE*/ΔSP will then have units of minutes^2. To fully normalise it, we divide by τ^2. Curve fitting gives a measure of performance against which almost all controllers should be assessed.

$$ITAE' = \frac{ITAE}{\Delta SP.\tau^2} = a_0 + a_1\left(\frac{\theta}{\tau}\right) + a_2\left(\frac{\theta}{\tau}\right)^2 \tag{13.15}$$

where

$$a_0 = 0.55 \qquad a_1 = 1.202 \qquad a_2 = 1.4250 \tag{13.16}$$

This is plotted as the coloured line in Figure 13.18. Completing the same exercise without the constraint on MV overshoot gives:

$$a_0 = 0 \qquad a_1 = 1.357 \qquad a_2 = 1.4164 \tag{13.17}$$

Also shown in Figure 13.18, as the dashed coloured line, this curve passes through the origin and meets the solid curve where θ/τ is around 1.8. As we saw in Chapter 3, the MV overshoot constraint has little impact on *ITAE* – even at very low θ/τ ratios. Also shown, as the solid black line, is the effect of switching to the more conventional PI-D algorithm (with the 15% limit on MV overshoot), described by:

$$a_0 = 0.55 \qquad a_1 = 0.600 \qquad a_2 = 0.9733 \tag{13.18}$$

This confirms what we know; this algorithm is theoretically preferred for SP changes. However, it will give a much poorer response to load changes than the I-PD algorithm. We would normally choose the I-PD algorithm, accepting the slightly higher *ITAE* during SP changes in order to benefit

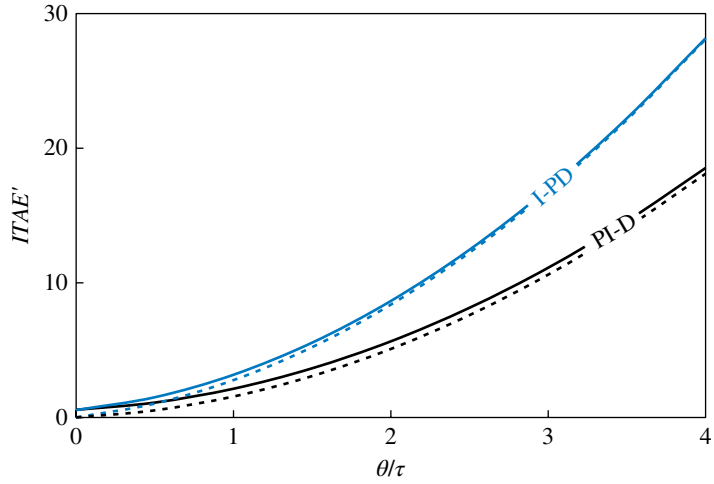

Figure 13.18 *Predicting minimum achievable ITAE*

from the much lower *ITAE* during more frequent load changes. Removing the MV overshoot limit with the P-ID controller gives coefficients:

$$a_0 = 0 \qquad a_1 = 0.564 \qquad a_2 = 0.9902 \tag{13.19}$$

This is shown as the dashed black line in Figure 13.18. It represents the theoretical minimum *ITAE* that can be achieved using PID algorithms typically found in the DCS. However, because of excessive MV movement and poor disturbance rejection, it is unlikely to represent an optimally designed controller. It is nevertheless a benchmark used by some commercially available tuning and monitoring tools.

The value of performance monitoring techniques lies more in detecting that something has changed. Monitoring over time would highlight issues such as a change in process dynamics or the development of a control valve fault. As an absolute measure of performance, the techniques have limited value. Firstly, they all require knowledge of the process dynamics. Deadtime is required to calculate the indices and what value they can attain is dependent on θ/τ. Once the dynamics are known then these can be used to obtain optimum tuning. Should this tuning be significantly different from that installed then, rather than simply use the dynamics to assess performance, an immediate improvement can be made by implementing the new tuning. Secondly, the techniques can only be practically applied to SP changes. Most controllers have to deal with far more load changes than SP changes. Reliably detecting the start of a load change and accurately assessing its size are both difficult. While potentially applicable to secondaries of cascades, these are likely to experience SP changes that are far too frequent for steady state to be reached between most of them. Finally we showed, in Chapters 3 and 4, that an optimally tuned controller is usually not defined as one with minimum variance but more often a compromise between this criterion and MV movement. Recording the sum of absolute changes in control output, and possibly also the number of reversals of output, give some indication of whether MV movement criteria have been met. Averaging level control can be monitored by recording how much of the permitted deviation is used. This might be by counting the number of times the maximum deviation is violated or by recording the closest approach.

Condition monitoring supplements performance monitoring and aims to detect and diagnose problems that might be affecting controller performance. Fundamental is oscillation detection. Oscillation is commonly associated with overly aggressive controller tuning but can also be caused by control valve problems such as stiction and hysteresis. We must be able therefore to not only reliably detect oscillation but also diagnose its cause. Detection is commonly based on analysis of the regularity of *zero-crossings*, i.e. occasions where the controller error changes sign. For this to be effective, spurious zero-crossings (such as those caused by measurement noise or normal process disturbances) must be eliminated from the analysis. We describe, in Chapter 15, filtering techniques commonly used to pre-process data to meet this objective. We also show in Chapter 14 how autocorrelation (or autocovariance) can be applied to reduce the impact of noise. Several criteria have been suggested to determine whether the zero-crossings occur at a regular interval. One [39] is to determine the mean and standard deviation of (typically) 30 consecutive intervals. The oscillation is then defined as regular if the standard deviation is less than one third of the mean. Another [40] determines the IAE between each zero-crossing. The result is compared to the suggested limit of $a.P_u/\pi$ – where a is typically 1% of instrument range and P_u is the ultimate period as described in Chapter 3. Oscillation is said to exist if the number of consecutive occasions that the limit is violated, within the suggested period of $50P_u$, exceeds the suggested value of 10.

There are mixed views on the value of such techniques. Oscillation of a noise-free measurement not subject to process disturbances will be visually apparent and there is little advantage in automatic detection. For the more difficult measurements the filter and zero-crossing interval parameters have to be individually adjusted to ensure reliable detection of regular oscillation. This can be time-consuming and not always successful. A more effective method is to analyse the frequency spectrum using the *discrete Fourier transform* described in Chapter 15. The presence of a dominant frequency is a good indicator of oscillation.

A common cause of oscillation is valve stiction. If suspected, this can be confirmed by increasing the controller gain (by steps of about 50%). With no stiction this would increase the amplitude of the oscillation – potentially to the point of instability. With stiction present the frequency of oscillation will increase with little change in amplitude. Alternatively a valve travel or *bump* test can be performed. With the controller in manual a series of small changes are made manually to the controller output. Stiction is confirmed if several changes are necessary for the valve to move. With more recently developed *smart* valves it is possible to historise the actual valve position in addition to the controller output. Plotting one against the other will, if stiction is present, result in a staircase shape. Several techniques have been proposed to automatically diagnose stiction by analysing the form of oscillation. One of the most basic is to identify periods where the difference between the largest and smallest PV is less than some defined threshold value. If, during the same period, the difference between the largest and smallest OP is greater than a defined threshold then stiction is deemed to have taken place. Stiction ratio can then be reported as the stiction period expressed as fraction of the total time.

Oscillation detected in one controller may be caused by a problem with another. To identify the root cause of the problem we have to switch to manual each potential source and determine whether the oscillation stops. We describe in Chapter 15 how frequency spectra can be cross-correlated to identify those controllers showing oscillation at the same frequency. This can be useful is identifying the source.

Following a range of academic research conducted recently, there is a growing range of proprietary software packages that generate a wide range of controller performance statistics and diagnostics. While they can be useful, they are no substitute for an expert control engineer systematically checking basic controller performance not only before embarking on an APC project but also as part of a regular review. Indeed, some packages can be counter-productive, distracting the engineer with too many statistics that have no real value.

13.6 Inferential Properties

While the effort in building inferentials is usually relatively small, the elapsed time can be very large. There are several questions that need to be addressed before the MPC project is started, these include:

- Do sufficient good quality data exist to support the development of the inferential?
- Should regression or a first-principle model be used?
- Should a specialist supplier be used?
- Should the inferential be built in the DCS or in a proprietary software package?
- Is the inferential sufficiently accurate?
- Should laboratory updating be applied and, if so, how?
- Should analyser updating be applied and, if so, how?
- If the inferential proves infeasible, what additional instrumentation should be included?

Although MPC implementers also offer inferentials they will deliver the best that their technology offers using the installed instrumentation. They will have little interest in working alongside a potential competitor supplying inferentials, nor will they wish to delay their development while additional data are collected or new instrumentation installed.

Work should therefore begin well in advance of any controller design work and certainly before any MPC implementation project is awarded. Many of the questions can be answered by the plant owner, supported by a specialist if required, by first attempting to develop regression type inferentials. It will quickly become apparent whether further data collection is required. For example it may be necessary to operate under different test run conditions with accurate time-stamping of samples taken at steady state. Automatic laboratory (or analyser) updating can be explored to see if this significantly reduces bias error or worsens random error. Accuracy can be compared to what is required. If sufficient accuracy cannot be achieved then a specialist can be brought in on a no-win no-fee basis to see if more accurate first-principle models can be developed. Regression models are easily implemented in the DCS using standard blocks. Other technologies require custom code or the use of a proprietary package. If neither approach is satisfactory then the installation of additional instrumentation can be explored – either to provide further inputs to an inferential or replace it altogether with an on-stream analyser. Such instrumentation can be long-delivery and its installation shutdown-critical. Step-testing cannot be completed until such instrumentation is in place.

13.7 Organisation

As with all projects the key to success of APC projects is commitment from senior management. Too often, process control is seen as a necessary evil – involving costly instrumentation, software and people. The only exposure most managers have had to process control is the theory they were taught at university (for which they probably still do not see the need) and a tour of the hardware installed in the control building. Many are still not convinced of its importance in maximising process profitability. It is perceived as an option; the process seemed to run just fine for years before it was installed, so why do we need it now?

A manager will authorise almost any necessary expense to reinstate a piece of failed equipment, without which the plant cannot operate. As a result process start-up may be advanced 24 hours and thus increase annual capacity by about 0.3%. Would the same manager authorise similar expenditure to improve by 15% the performance of MPC that can achieve a 2% increase in capacity utilisation? Both have the same effect on production.

For a project to have long term success a management culture is needed which asks 'Why is APC not installed?' as opposed to 'Why should APC be installed?' While it may be possible to convince management to sanction the project, the risk is that this commitment is short term. When there is pressure to relocate key personnel to other areas of the business, this may be done to the detriment of APC. Performance will then slowly degrade and a major initiative, probably as costly as the original project, will be required to re-establish the capture of the benefits. Often such an initiative will not be forthcoming until there is a change in management.

A management 'champion' will ensure everything necessary for project success is put in place. There will a strong commitment to ongoing projects, executed as part of an agreed master plan. Approval of expenditure will be rapid. Staff of the highest competency will be assigned to the project(s) in sufficient numbers. Work outside the control of the APC team, such as instrumentation improvements, will be given the correct priority. The importance placed on APC will become apparent to the vendor who will then assign better staff to the project and pay greater attention to their client. The more senior the manager, the more successful will be the project. The benefits captured by APC are often said to be proportional to the salary of the manager sponsoring the project. Often the champion is only in place by chance and then may soon be replaced by someone less enthusiastic. Commitment to APC is rarely a criterion used in appointing a manager.

The implementation team can do much to nurture what interest there is. To maintain the momentum, rather than seeking approval separately for individual projects, commitment should be sought for all the projects conceived in the master plan. While the sanction process might be lengthier, it is likely the larger budget requires that a wider group of staff and more senior management are involved – further raising the profile. Given that implementation would then take several years, it would more likely survive the periods when less enthusiastic managers are in place.

Ongoing benefits are not as obvious as the same benefits first captured on commissioning APC. Even less obvious is the slow decline that neglect will cause. A new manager may not be aware of what was achieved by the project. Many of the monitoring tools described in this chapter offer the opportunity to regularly engage with senior management to help maintain a high profile for APC and to provide continuity during changes in management personnel.

It is a common problem that management mistakenly assume that less expertise is required for ongoing support compared to what was necessary for implementation. While it is possible to contract the APC vendor to take a major involvement in implementation, ongoing support requires in-house expertise. This is illustrated by Figure 13.19. The net benefits captured by a project implemented and supported by a good in-house project team will typically follow the solid black line. However, if the same level of expertise is not retained for ongoing support, the benefits will decline substantially – as shown by the dashed line. A more expert team, for example one supplemented throughout implementation and support by very experienced staff, would commission controllers more rapidly and would continue to identify ongoing opportunities to capture further benefits – as shown by the coloured line.

The performance of basic instrumentation, the accuracy of inferentials, the availability of on-stream analysers and the variability of the MPC dynamics all need close monitoring. There needs to be a check at least daily that each MPC has not been over-constrained, is using the correct objectives coefficients and is driving the operation against the correct constraints. There needs to be frequent liaison between all the groups that can influence the success of the controller – including process supervision, process technical support, planning and economics section, instrument and system support personnel and process operators.

Regular (e.g. monthly) meetings, chaired by the APC engineer, should be attended by the plant manager, process technical support, process economics specialist, instrument technician and system support. The main agenda items include APC performance for the last month, problems encountered, solutions developed, forthcoming changes and an agreed action plan (including who, what and when).

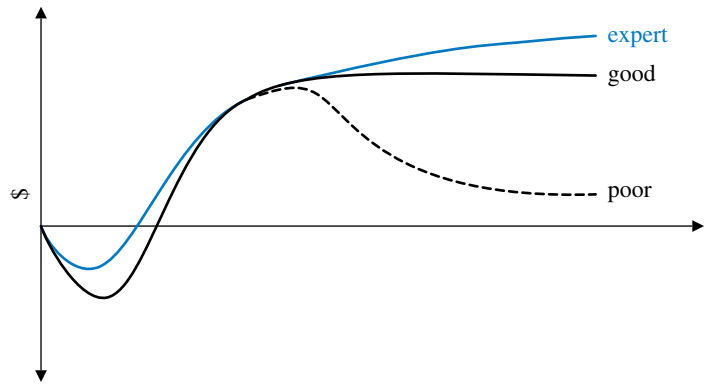

Figure 13.19 *Impact of level of post-project support*

An experienced APC engineer should be involved in the approval of the design of all proposed process modifications. Additions agreed at the process design stage can be implemented for almost no incremental cost, compared to the possibly unjustifiable costs of retrofitting the change. Examples include ensuring sufficient tray temperature thermowells are installed on distillation columns, ensuring sufficient meter runs and orifice flanges are installed for the later addition of flow meters and ensuring sufficient space is reserved for the future installation of an analyser house. Such involvement will also help avoid many poor process design practices that later cause process control difficulties. These include wrongly placed and wrongly ranged level gauges where surge capacity can be used, omission of instrumentation important to inferentials, selection of control strategies known to give problems with inverse response or have other dynamic problems.

While recruiting good control engineers is possible there is an ongoing need to train existing staff. Projects present an ideal opportunity for on-the-job training. Trainees should be selected firstly on their level of enthusiasm for the subject and, because they need to liaise with almost every part of the organisation, on their interpersonal skills. They need a strong process background acquired either through education (e.g. chemical engineering graduates) or though experience (e.g. former process operators). They will need a general appreciation of process economics.

New engineers should attend training courses in DCS configuration, usually provided by the DCS vendor, and in the chosen MPC package, usually by its developer. What is often overlooked is training in the areas covered by this book. There are many courses offered in the academic world, the majority of which are unashamedly highly theoretical. Such courses are probably the biggest cause of potentially very competent control engineers choosing another branch of the engineering profession. A small minority of the academic institutions are beginning to appreciate the difference between theory and practice, but often the lecturers have not been in a position to accumulate the practical expertise to pass on.

In assigning a new owner's engineer to a project team it is important that the APC vendor understands what is required in terms of training. Most vendors do not see training client staff as a threat to their business and most will enthusiastically take on the training role. They do, however, need to factor in what impact it might have on their man-hour budget and schedule.

Once trained, retaining the expertise presents another problem. APC vendors are always on the lookout for good staff. A newly trained engineer will be looking for the next challenge; if all that is offered is ongoing support, rather than another project, working permanently on projects with an APC vendor might appear attractive. Much depends on the value the employer places on technical expertise. Too often

those following technical careers are perceived as doing so because they would fail as managers. Many companies prefer to train generalists rather than specialists. Ideally a career in APC in larger companies should offer development into some centralised engineering/consultancy role that carries the same kudos as the equivalent position in the management hierarchy.

Other moves which help retain expertise is to have an approved master plan in place so that staff can see firstly a commitment to the technology by the company and secondly a role for themselves on future projects. Promoting APC engineers into management positions can demonstrate that the experience gained is valued by the company; few technical positions in a manufacturing company require as much understanding of the business as that developed by APC engineers.

Rotating staff through the APC group is beneficial on two counts; it imports knowledge of the process operation or its control systems and it exports APC supporters into other parts of the organisation. Many successful APC projects are on plants now managed by former APC engineers. Some of the most successful APC engineers were once process operators. Similarly exchange of experienced personnel with other sites, or even with vendors, will help develop expertise. Importantly movement demonstrates that joining the APC group opens up career opportunities rather than closes them down. Successful rotation routes are shown in Figure 13.20.

The debate in many organisations is where to locate the process control team. Their function is astride the interface between process technical services and instrumentation support, so arguments are made for them to reside in one group or the other. The group also has strong links with operations department and the planning and economics group, so there is logic in locating them in either of these. In fact, all of these options have been explored somewhere and each has succeeded and each has failed. The important consideration is not where but who. A well-chosen, well-trained engineer reporting to a manager well versed in APC will be effective in just about any part of the organisation.

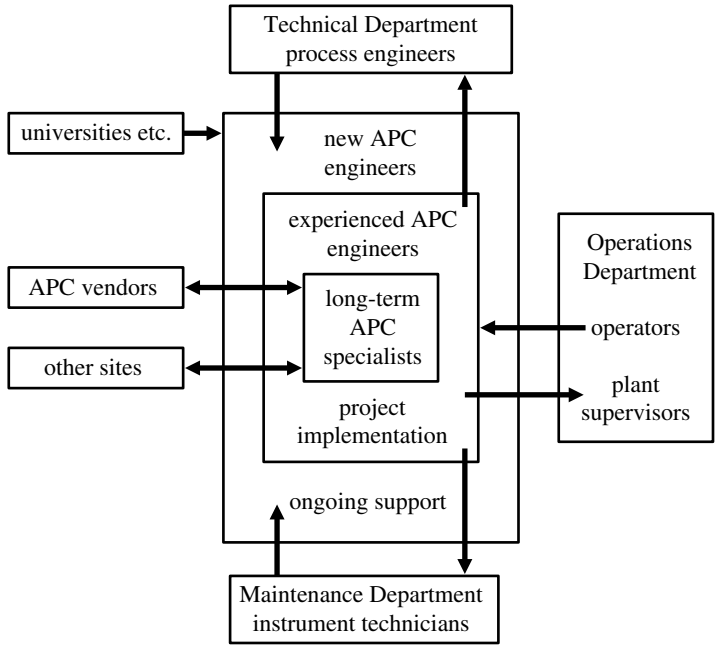

Figure 13.20 *Staff rotation*

The structure which does have several reported failures is the plant-centric approach. Here, all the staff associated with a particular plant are located in the same section reporting to the plant manager. Often the plant may not be complex enough to require the full-time attention of one APC engineer and so the role will be diluted by other responsibilities. Further not all plant managers are equally enthusiastic about APC. The APC engineer, who in a technology-centric structure would otherwise work on other processes for multiple managers, can become disillusioned with the profession or employer.

MPC is a complex technology not always readily understood by all console operators. Technical support is usually only on site for 25% of the time, for the other 75% the console operator has the greatest influence over APC utilisation. If not properly trained, one operator can take an action which loses in one shift the profit that the controller has made in one year. There needs to be at least one operator per shift fully conversant with the APC. This is best achieved by assigning, to the implementation team, a respected leading operator who will be seen by the other operators as their representative. As a result there will be much greater confidence in the APC. Plus the lead operator, when returning to the normal shift position, will be an invaluable source or expertise outside of office hours.

APC expertise within a company can often be overstretched in trying to support existing APC applications while also being involved in a major implementation project. APC is treated quite differently from other additions to the process. For example a new compressor, once accepted as operating correctly, becomes the responsibility of the plant manager who will coordinate any necessary maintenance and organise any additional training prompted by operator misunderstanding. Only in the event of more complex problems will the specialist be again involved. There is no reason why much of APC monitoring should not be delegated in this way. Once the controls are accepted, the plant manager's role should include routine checks that it is not being over-constrained and is operating against the right constraints. Tools developed for monitoring on-stream analyser performance can be passed to the analyser technicians and/or engineer. Similarly, inferential performance monitoring can be passed to the process support engineer, as can the responsibility for reporting APC performance. The APC engineer would only be involved in the event of problems that fall outside of such routine support.

Prior to commissioning MPC there is a need to involve the planning and economics group in a review of the process gain matrix, the economics used and the result of any off-line MPC simulations. While at first they might feel that they might struggle to understand the technology, it should quickly become apparent that the steady state part of the controller is very similar to LP-based planning tools. They will recognise process gains as *vectors*, objective coefficients as *reduced costs*, etc. The operating constraints as understood by the planning group will be conservative and so any operating plan is likely to be sub-optimal. Recognising that some of these constraints can be exceeded if required will influence the way they specify the operating plan. For example an instruction to run at a particular feed rate might be replaced with 'maximise feed rate, but not more than …'. They should also be given access to the offline version of MPC so that they can explore the impact of changing economics before passing them to operations department. They should also be able to view the on-line version to check that new strategies are being adhered to.

13.8 Vendor Selection

There is a growing number of APC technology suppliers and implementers. Some implement only technology that they have developed. Others have exclusive licences to work with just one technology developed by others. And there are others that theoretically offer a choice of technologies – although in practice they will often favour one over the others. The situation is further complicated by specialist suppliers. For example there are those that specialise in inferential properties, those that have developed CLRTO packages, those that offer process-specific technology, etc.

Choice of vendor and implementer can be bewildering to an engineer, especially for one doing it for the first time. The first step is the production of an *invitation to bid (ITB)*. In addition to the plant owner's normal commercial terms, this should include:

- brief description of the process; simplified process flow-sheets showing the basic controls will help the bidder assess the number of MVs and hence the work involved in step-testing and commissioning
- brief description of the control system, including its network(s), modules and how they are split between processes
- description of data acquisition system(s) and availability of historical data
- history of any previous APC implementations
- the key results of the benefits study listing, in engineering units, the anticipated process improvements
- list of any existing inferentials and their accuracy
- list of personnel that the client plans to assign to the team, their experience, what their role is and how the implementer is expected to involve them
- any specialist organisations that the successful bidder will be expected to work with and whether they will be subcontracted to the implementer or contracted to the plant owner
- list of documentation expected and what language should be used
- what performance guarantees are required
- any special procedures that the vendor must follow, for example safety reviews
- pricing basis required, i.e. lump sum or reimbursable, travel and living expenses included or rechargeable, staged payments, etc.

The ITB should also describe what is required of the bidder, including:

- list of reference sites with contact details for similar work completed by the bidder for other clients
- list (with résumés) of those staff short-listed for each of the project roles
- proposed project schedule bar-chart
- description of involvement of plant owner's personnel
- contents of any training courses offered
- examples of specimen documentation

Once the ITB has been issued, but before the bid due date, meetings should be held with all the bidders. These serve two purposes – an opportunity to clarify the ITB as required and assess the bidders' capabilities. Ideally this should be in their headquarters since this usually gives an opportunity to meet with a wide range of the bidders' technical staff and see demonstrations of their products. However, many of the smaller implementation companies do not have offices; their personnel largely work from home or on site. In which case the meeting can be at any convenient location, providing that the bidder brings to the meeting likely members of the project team.

The plant owner should take the opportunity to visit a selection of reference sites. This helps a little with vendor selection but the vendor is unlikely to suggest a site where his reputation is poor. It is more an opportunity to benefit from the experience of others who have completed similar projects.

Bid analysis is important to check that each bidder has complied with the requirements; clarification should be sought where necessary. How much it contributes to vendor selection can vary greatly. Sometimes, from the contact the plant owner has had with each bidder, selection may well have been made before the bids were submitted. Clarifying the bid is only necessary since it will become part of the contract.

On other occasions, vendor selection can become quite complex. This can arise where one bidder has strengths not matched by others but has weaknesses compared to the others. If a large number of the owner's personnel are involved in the bid selection there can often be very different views on which is the preferred bidder. Under these circumstances a more methodical approach is required.

Kepner-Trago analysis is a decision support methodology. It is occasionally discredited and can have the reputation that it only confirms a decision already made. This can arise if it is not applied impartially. The technique is to first brainstorm, involving all the staff involved in the decision, and list all the criteria on which selection should be based. For each criterion, a numerical value should be placed on its importance. These values should not be just arbitrary weighting factors. They should as far as possible represent the financial impact of meeting, or failing to meet, each criterion. It is useful to generate this table before finalising the ITB, so that any additional information required from the bidders can be included. Analysis of the bids then includes scoring each bidder against each criterion and totalling the financial impact. Table 13.1 shows a much abbreviated example for one bidder.

Once the preliminary analysis is complete then, to aid selection:

- Delete from all of the analyses any issue on which the bidders score equally, so the focus is on differences.
- Delete any issue valued so low that it cannot have any impact on the decision, so the focus is on important differences.
- Explore how sensitive the result is to the ratings given to determine if the most pessimistic or optimistic view would alter the result.
- Discuss with the bidder any items where he might be able to improve his rating.

The fees quoted can vary widely from bid to bid. Bidders may be bidding tactically, e.g. to win a client already working with a competitor or to add an important site to its references. Bidders can also inflate their quotation if not enthusiastic about winning the work but think it may harm their relationship with the owner if they decline to bid. However, it is important to reconcile any major differences to ensure that the bidder has properly understood the scope of work.

Clearly cost is an issue in selecting the bid but needs to be balanced by consideration of the benefits. The result of the Kepner-Trago analysis permits exactly this, financially justifying the decision to select a higher quotation.

No bid should be rejected until negotiations are completed with the leading bidder. Indeed, if a single tender approach has been adopted, the sole bidder should still be allowed to believe that he is bidding competitively. There are a wide range of techniques for doing this, for example revising the bid submission date at the request of 'others'. While negotiation on price is a possibility, more important is to get the bidder to agree to assign the best engineers to the project. The owner should identify what makes the project attractive

Table 13.1 *Kepner-Trago analysis*

Key issue	Potential benefit	$k	Rating (%)	Value
Training of owner's staff	50% cost saving on next project	150	50	75
Inferential technology	cost of analyser + 9 months benefits	175	70	123
Lead engineer quality	20% of total benefits for 3 years	600	85	510
Location of support	3 faults/year fixed 1 day earlier	25	100	25
Overall		950	77	733

to the bidder. For example, the bidder might see particular value in being able to bring future potential clients to the site once the project is complete and therefore be prepared to offer something in return. The owner is in an even stronger negotiating position if previous contracts were awarded to the bidder's competitor. A reference site where the bidder has displaced its competition is a valuable marketing advantage.

13.9 Safety in APC Design

While safety should be uppermost in any design decisions there is wide variation in how this is managed. The attitude of some engineers is that safety should be handled by the basic controls and, if anything, the APC will directionally improve safety. However, there are many examples of an APC application causing a major process upset. Further it is often the case that the basic controls are modified to support the addition of APC.

Some operating companies have adopted a formal approach to APC safety, treating it in much the same way as any other plant modification. The well-documented *Hazard and Operability (HAZOP)* approach is generally too rigorous for APC, since it is difficult to prevent it from becoming a full process review. However, if HAZOP is being applied to a major process revamp, APC should be included in the review. The *Control, Hazards and Operability (CHAZOP)* study is more suited to identifying the risks associated with the control system itself. *Failure Mode Event Analysis (FMEA)* has been used successfully although, in addition to the risks arising from APC, it will identify problems with the basic controls already in place. The key to success is to ensure the review meeting has a strong leader whose role is to prevent the discussion moving outside the area of process control and also ensure every point raised is properly followed up.

Some companies have extended their permit-to-work system to include APC. Such permits require the signature of key personnel such as the plant manager, the head of the APC team, etc. Permits are required for step-testing as well as commissioning.

Once safety procedures are agreed it is also necessary to agree criteria about when they are applied. What level of modification to a control strategy requires a formal review and/or a permit? Most would allow changes to operator graphics without a permit. Many allow changes to tuning. But would a change to a different version of the PID algorithm justify the paperwork? Would the addition of signal conditioning, etc.?

13.10 Alarms

A problem that can arise with the use of DCS is a proliferation of alarms. Each controller can be configured to give alarms for HI/LO PV, high rate of change and excessive deviation from SP. Other messages can be generated to draw attention to the change of auto/man mode, configurations changes etc. Left unchecked, the large number of alarms can cause the process operator to miss those that are particularly serious. Several major incidents are on record where the deluge of alarms was later identified as a contributory factor.

While not strictly in the area of process control, a project and its associated budget do give the opportunity to address the issue. Experiences of those having performed such exercises are remarkably consistent:

- It is essential that operations department are strongly committed to the work; indeed without this it is probably not worth progressing.
- A strong project team, including a very experienced process operator and process engineer, is required. The design, support and monitoring of any resulting alarm management system usually fall to the control engineer.

- An alarm reduction study, possibly assisted by external consultants, should be progressed before the decision made to install any alarm management package.
- Resolving the problem is not a one-off piece of work; there is a need for a regular review of all alarm actuation and revision of alarm management configuration.

The team are set the objective of meeting the guidelines [41] published by Engineering Equipment & Materials Users' Association. These recommend the following upper limits per operator console:

- No more than 10 *standing alarms*, i.e. alarms which have been acknowledged
- No more than 10 *background alarms* per hour, i.e. alarms for information purposes that do not require urgent attention
- No more than 10 alarms in the first 10 minutes after a major process problem develops

There are a number of commercial alarm management systems available. Some features are often already built into the DCS as standard. Others require the addition of a package – maybe from a third party. These are not a substitute for an ongoing review of alarm statistics, but they do add useful functionality. They can be particularly useful in identifying repeating *nuisance* and long-standing alarms.

- Alarms can be categorised so that the greatest attention is drawn to the most important alarms.
- Alarms can be suppressed; this can be applied to alarms known to be the result of an upset that has already generated a *first-up* alarm. Or it can be used in predictable situations such as start-ups and routine shutdowns.
- Individual alarms can be configured with a deadband or an off-delay timer to stop alarm *chattering*. An on-delay timer will stop *fleeting* alarms.
- Retrospective analysis of incidents is supported by tools that enable the alarm database to be searched for unnecessary duplicate alarms, multiple occurrences of the same alarm, exact timing and sequence of alarms, alarms associated with the same equipment or plant area, etc.

14

Statistical Methods

Statistical methods play a significant role in benefit estimation, reconciling process data, the development of inferentials, model identification and monitoring the performance of process control. Most engineers apply these methods quite correctly, often using proprietary software or spreadsheet functions, without necessarily understanding their background. However, there are often occasions where the methods are misapplied; or licence fees are unnecessarily incurred for software that could easily be replicated by the control engineer using a spreadsheet package.

A key objective of this book is to avoid unnecessarily taking the reader into theoretical detail. And indeed it is expected that most users of the techniques described need never read this chapter. It is no accident that it is included towards the end of this book. However, the reader is encouraged to brave the mathematics involved. A deeper understanding of the available techniques should at least be of interest and potentially of great value in better understanding services and products that might be offered to the control engineer.

14.1 Central Limit Theorem

Proof of the Central Limit Theorem involves mathematics beyond the scope of this book. Nevertheless its conclusion is of great importance – particularly in supporting the assumption that the variation in key process performance parameters follows a *normal* or *Gaussian* distribution. Fundamentally, the theorem states that summing a large number of independent values, selected randomly, will result in a total that is normally distributed – no matter what the shape of the distribution of the values selected. In effect a dependent process parameter, such as product composition, is determined by the combination of a large number of independent variables, such as flows, pressures and temperatures. Each of these independent variables does not have to be normally distributed but the dependent variable will be.

As a simple example consider the results of throwing a single unbiased six-sided dice. There is an equal probability (of 1/6) of throwing any number from 1 to 6. Throwing the dice enough times would result in a *uniform* distribution of the score (x) with a *mean* (μ) and *variance* (σ^2) given by

$$\mu = \frac{\sum_{i=1}^{6} x_i}{6} = 3.5 \tag{14.1}$$

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

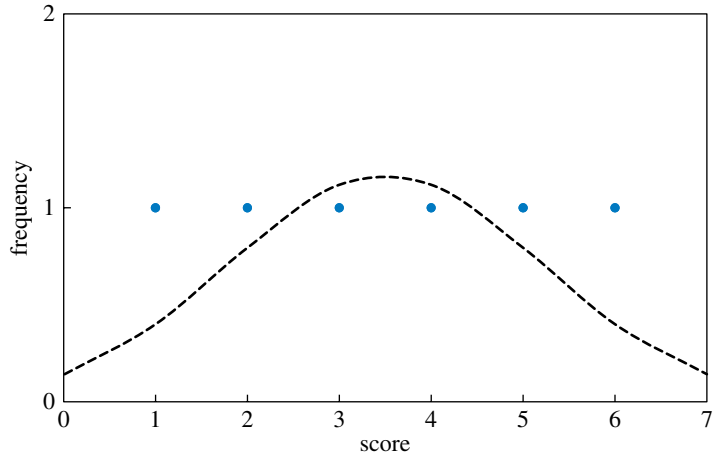

Figure 14.1 *Frequency distribution for a single dice*

$$\sigma^2 = \frac{\sum_{i=1}^{6}(x_i - \mu)^2}{6} = \frac{35}{12} = 2.9 \tag{14.2}$$

Figure 14.1 shows the normal distribution, with the same mean and variance, superimposed on the true distribution. Clearly the distributions are very different. However, if we sum two values each selected from the uniform distribution, the equivalent of throwing two dice, we obtain the distribution shown in Figure 14.2. The mean and variance of this distribution are obtained by summing the means and variances of the source distributions, giving values of 7 and 5.8 respectively. Figure 14.2 shows that the now triangular distribution is closer to the normal distribution. Closer still is the distribution for throwing five dice as shown in Figure 14.3. It will have a mean of 17.5 and a variance of 14.6.

We have demonstrated how summing values selected randomly from a uniform distribution will generate a normal distribution, but the Central Limit Theorem goes further than this. Consider a six-sided dice with

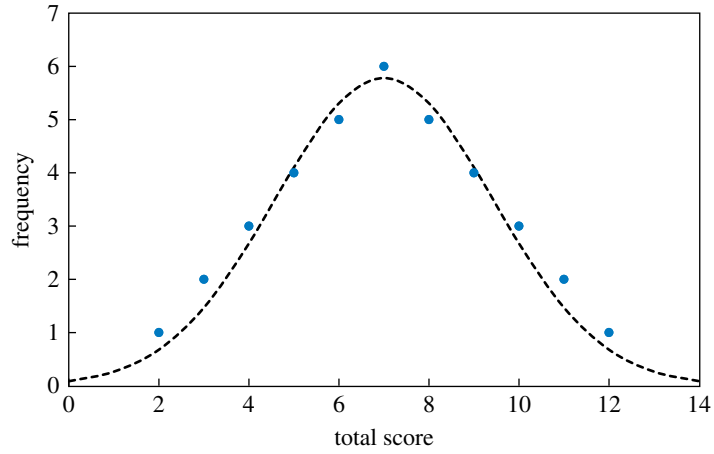

Figure 14.2 *Frequency distribution for two dice*

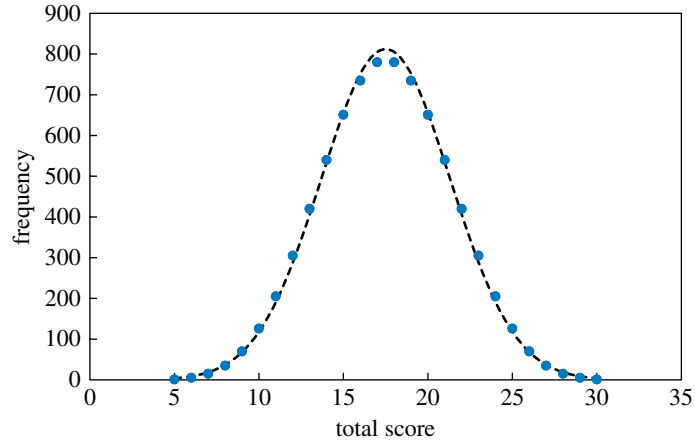

Figure 14.3 Frequency distribution for five dice

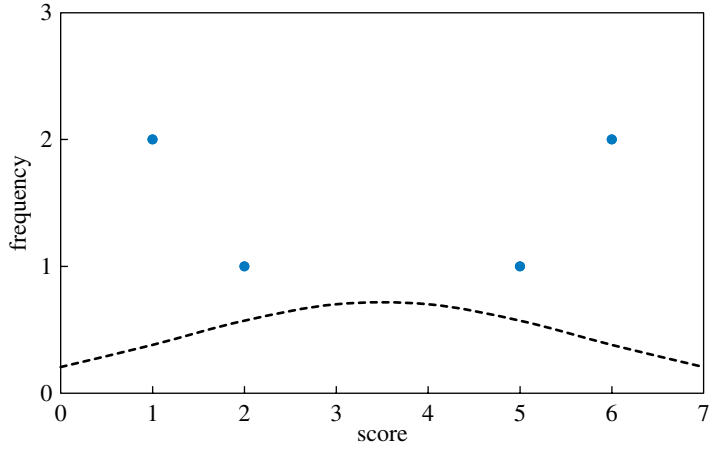

Figure 14.4 Frequency distribution for a single modified dice

the sides having values of 1, 1, 2, 5, 6 and 6. Throwing such a dice will generate the distribution, shown in Figure 14.4, which is far from uniform. Throwing a large enough number of such dice will generate a normal distribution. For example, with seven such dice we begin to approach it, as shown in Figure 14.5.

14.2 Generating a Normal Distribution

With dice, since only integer scores are possible, the mean and standard deviation must be calculated from discrete values. For continuous functions a slightly different approach is required. Consider a number randomly generated from a uniform distribution in the range 0 to 1. Frequency, on the vertical axis, is replaced with *probability density*. Since we are certain to generate a value in this range, the area under the distribution curve must be unity – as shown in Figure 14.6.

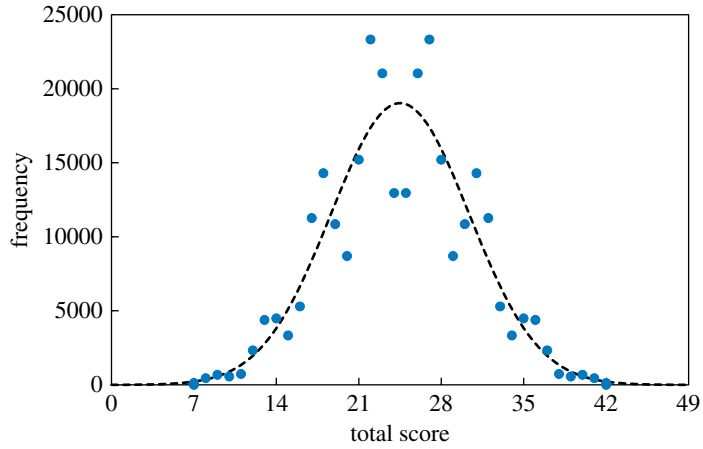

Figure 14.5 *Frequency distribution for seven modified dice*

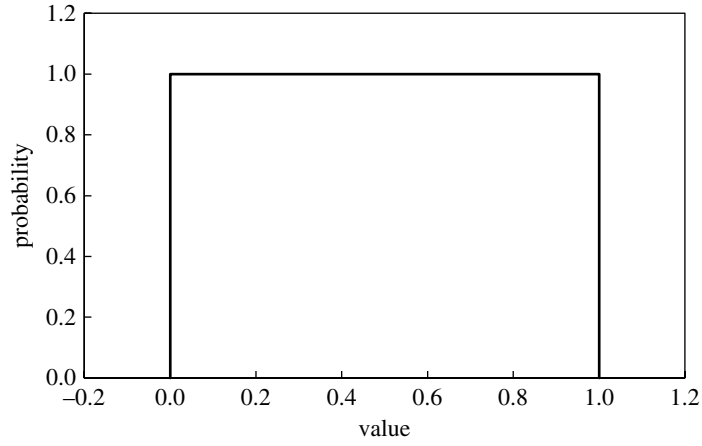

Figure 14.6 *Probability density for uniform distribution*

The mean of the distribution is 0.5 and the variance given by

$$\sigma^2 = \int_0^1 (x-0.5)^2 .dx = \int_0^1 \left(x^2 - x + 0.25\right).dx = \left[\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right]_0^1 = \frac{1}{12} \tag{14.3}$$

As we saw in Section 14.1, if we add together 12 values (x_1 to x_{12}) randomly selected from the same uniform distribution, we would obtain a value chosen from a distribution very close to normal. It would have a mean of 6, a variance of 1 and thus a standard deviation of 1. This gives us a method of generating a value (x), for example for use in *Monte Carlo simulation*, from a normal distribution required to have a mean of μ and a standard deviation of σ.

$$x = \mu + \sigma\left[\sum_{i=1}^{12} x_i - 6\right] \tag{14.4}$$

The normal distribution curve itself can be drawn by applying the formula for the *probability density function* (P_x) for a value (x).

$$P_x = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \tag{14.5}$$

Often this equation is converted to the *standard normal function* with a mean of 0 and a standard deviation of 1. Sometimes described as N(0,1), it will have the formula

$$P_z = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-z^2}{2}\right] \tag{14.6}$$

In addition to the normal (Gaussian) distribution, there are around 500 others published. Many have very specific uses and are unlikely ever to be required by a control engineer. Later in this chapter, we describe the few that are occasionally of value.

14.3 Quantile Plots

Quantile plots (or *Q-Q plots*) are used to compare the distributions of two sets of values. The plot requires each set of values to be ranked in increasing order. While changing the sequence of the values has no effect on mean or standard deviation, there is no guarantee that values in one set will be sorted into the same order as the other. For this reason the technique cannot be used to determine whether the two variables are correlated. For example, it cannot be used to validate an inferential property against its corresponding laboratory result. The use that the control engineer might make of it is to confirm that data collected for statistical analysis, such as benefit studies, are normally distributed. This is achieved by plotting the expected values, if the data were normally distributed, against the corresponding observed values.

The reader is likely to be familiar with the term *quartiles* – the three points that divide the ranked data into four equal groups. Quantiles are the points which divide the data into n equal groups, where n is the number of data points and so each group will contain one data point. The data points are first ranked, where k is the ranking of each point. The quantile (α) for each point is then determined from

$$\alpha = \frac{k-a}{n+1-2a} \tag{14.7}$$

The term a is chosen (usually in the range 0.0 to 0.5) although, if n is large, the choice has a negligible impact on the conclusion. Depending on its choice each quantile is therefore calculated from

$$a = 0.0 \qquad \alpha = \frac{k}{n+1} \tag{14.8}$$

$$a = 0.5 \qquad \alpha = \frac{k-0.5}{n} \tag{14.9}$$

The expected value (x) corresponding to each quantile is calculated using the mean (μ) and standard deviation (σ) of the observed data. If we are checking whether the distribution is normal we use the *cumulative distribution function* derived by integrating the probability density function as described by Equation (14.5) and solve for x.

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-(x-\mu)^2}{2\sigma^2}}.dx = \alpha \tag{14.10}$$

This is solved by iteration; fortunately most spreadsheet packages include this as a standard function. We can choose any value for the mean and standard deviation (e.g. $\mu = 0$, $\sigma = 1$). If the data are normally distributed then the plot of actual values against those expected will be a straight line. For the line to pass through the origin and have a slope of 1 then we must use the mean and standard deviation derived from the data. To highlight any deviation from normal distribution we can add a reference line – commonly drawn through the first and third quartiles. It is this method that was used to prepare the quantile plots included in Chapter 13.

Rather than comparing a set of values to an expected distribution, we may wish to compare two sets of values to determine whether they have the same distribution. For example, we might want to compare two contending inferentials. The data sets are the prediction errors for each. Both data sets need to contain the same number of values. If this is not the case then the quantiles in one set, which are missing from the other, need to be added by interpolation – and vice versa. To do so, we first identify the ranking (k and k+1) of the two values to be used for interpolation. If we have used Equation (14.8) to determine α then

$$k = \text{int}\left(\alpha_{missing}\left(n+1\right)\right) \tag{14.11}$$

We then determine the weighting factor (w) that should be applied

$$w = \alpha_{missing}\left(n+1\right) - k \tag{14.12}$$

The interpolated value (x) is then given by

$$x = \left(1-w\right)x_k + wx_{k+1} \tag{14.13}$$

The resulting quantile plot will highlight any regions where the prediction from one inferential is consistently different from the other. This can be used to help identify the cause of the error and explore improvements.

14.4 Calculating Standard Deviation

The standard deviation (σ_p) of the whole *population* of N values, if the *population mean* is μ, is given by

$$\sigma_p^2 = \frac{\sum_{i=1}^{N}\left(x_i - \mu\right)^2}{N} \tag{14.14}$$

However, in practice we select a sample from the population. For example, when executing process control benefits studies we select a period for analysis comprising n data points. It would not normally be practical to analyse all the data collected since the process was first commissioned. Those performing such analysis will likely have noticed that, to account for this, μ in Equation (14.14) is replaced by the sample mean and N is replaced by $n - 1$. The following explains why.

From the data points collected in the period we estimate the sample mean.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{14.15}$$

Applying Equation (14.14) to a sample of the population will underestimate the true standard deviation. This is because the sum of the squared deviations of a set of values from their sample mean ($\bar{x}$) will always be less than the sum of the squared deviations from a different value, such as the population mean (μ). To understand this, consider the trivial example where we have a sample of two data points with values 1 and 5. Their mean is 3 and the sum of the deviations from the mean is 8 ($2^2 + 2^2$). Instead of using the mean, we choose to use a value of 4; the sum of the deviations will then be 10 ($3^2 + 1^2$). The nonlinearity, caused by squaring, results in the increase in squared deviation in one direction being greater than the decrease in the other.

We do not know the mean of the whole population (μ). Applying Equation (14.15) to different samples selected from the population will give a number of possible estimates of the true mean. These estimates will have a mean of μ. Similarly we do not know the standard deviation of the whole population. Imagine the sample mean being determined by, before summing all the data points, dividing each by n. The standard deviation of the resulting values will therefore be n times smaller, i.e. σ_p/n, giving a variance of $(\sigma_p/n)^2$. Variances are additive, so the sum of the n values (which will now be the sample mean) will have a variance n times larger, i.e. σ_p^2/n.

The variance of the sample (σ^2) is given by

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \tag{14.16}$$

The variance of the population will be the variance of the sample plus the variance of the sample mean.

$$\sigma_p^2 = \sigma^2 + \frac{\sigma_p^2}{n} \quad \text{or} \quad \sigma_p^2 = \frac{n}{n-1}\sigma^2 \tag{14.17}$$

Substituting for σ^2 from Equation (14.16) gives

$$\sigma_p^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \tag{14.18}$$

We use Equation (14.18) to generate an *unbiased variance*. This technique is known as *Bessel's correction*. In practice, if the number of data points is sufficiently large, the error introduced is small. For example, with a value of 50 for n, the effect on σ^2 will be to change it by about 2%, with a change in σ of less than 1%.

To remove the need to first calculate the sample mean, Equation (14.18) can be rewritten as

$$\sigma_p^2 = \frac{\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2}{n-1} = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} \tag{14.19}$$

14.5 Skewness and Kurtosis

Two related parameters, known as *skewness* and *kurtosis*, can be used to determine how close the distribution is to normal. The reader, familiar with these parameters, may be used to simpler formulae. Like that developed in Section 14.4, for standard deviation, these here have been modified to apply to a sample taken from the whole population of data points.

$$skewness = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{\sigma_p^3} \tag{14.20}$$

A normal distribution is symmetrical about the mean. If the skewness is greater than zero then the distribution is skewed to the right, i.e. the higher values are further from the mean than the lower values. Typically, provided skewness is between −0.5 and +0.5, we can treat the distribution as normal.

$$kurtosis = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{\sigma_p^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{14.21}$$

Kurtosis is a measure of how flat or peaked is the distribution. Strictly Equation (14.21) gives the *excess kurtosis*. The kurtosis of a normal distribution is 3; the excess kurtosis will be 0. If excess kurtosis is positive then the distribution is *leptokurtic*, i.e. is peaked with long tails. If negative then it is *platykurtic*, i.e. flat with short tails. Again, if kurtosis is between −0.5 and +0.5, we would treat the distribution as normal.

14.6 Correlation

The calculation of *covariance* (σ_{xy}) is a useful first step in determining whether there is any correlation between two variables (x and y). It is defined as

$$\sigma_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{14.22}$$

If there is no correlation between x and y then the covariance will be close to 0. However, if y tends to be above its mean when x is above its mean then it will have a positive value. Similarly if one variable tends to be below its mean when the other is above then it will have a negative value. This can be a useful step in dynamic model identification. Determining the covariance between PV and MV will tell us whether the process gain is positive or negative.

The variance of a value derived from two or more measurements, for example a change in inventory determined by subtracting the flow out from the flow into a process, is normally determined by summing the variances of the two measurements. This is correct if the error in one measurement is not influenced by the other, i.e. they are truly independent variables. However, if they are correlated, calculation of the combined variance must take account of the covariances.

$$\sigma_{x-y}^2 = \left(\sigma_x^2 - \sigma_{xy}\right) + \left(\sigma_y^2 - \sigma_{yx}\right) = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy} \tag{14.23}$$

We can see from Equation (14.22) that σ_{yx} is the same as σ_{xy}. By convention we write it in this way because we subtract, from the variance of y, its covariance with respect to x.

Measurement errors might be correlated if, for example, they have a common cause. For example, a change in fluid properties might affect both flow measurements similarly. Note that the variance of the sum of correlated measurements is determined by modifying Equation (14.23) so that the covariances are added rather than subtracted. For example, estimating the feed to a process by adding correlated measurements of its three product flows would have a variance calculated from Equation (14.24)

$$\begin{aligned}\sigma_{x+y+z}^2 &= \left(\sigma_x^2 + \sigma_{xy} + \sigma_{xz}\right) + \left(\sigma_y^2 + \sigma_{yx} + \sigma_{yz}\right) + \left(\sigma_z^2 + \sigma_{zx} + \sigma_{zy}\right) \\ &= \sigma_x^2 + \sigma_y^2 + \sigma_z^2 + 2\left(\sigma_{xy} + \sigma_{xz} + \sigma_{yz}\right)\end{aligned} \tag{14.24}$$

The limitation of covariance is that it is difficult to determine the significance of its value, since it depends on the magnitude of the variables. When x and y are identical the covariance will reach its maximum possible value. This will be the variance of x (which is also the product of the standard deviations of x and y). Dividing the covariance by the standard deviations of x and y gives the dimensionless *Pearson coefficient* (R).

$$R = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} \tag{14.25}$$

To avoid first having to calculate $\bar{x}$ and $\bar{y}$, this can be rewritten as

$$R = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right)\left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}} \tag{14.26}$$

Pearson R will be in the range −1 to +1. If there is an exact correlation (such that $y_i = mx_i + c$) then R will be +1 if m is positive, and −1 if m is negative. Often R^2 is used to remove the sign. A value for R of 0 means that there is no relationship between x and y. It is important to appreciate that a nonzero value of R does not indicate that the values of x are necessarily close to the corresponding values of y. It is therefore of limited value in assessing whether an inferential property closely matches the corresponding laboratory result.

14.7 Confidence Interval

The confidence interval is used to indicate the reliability of an estimate. For example, a 95% confidence interval means that there is a probability of 0.95 that the true value is in the quoted range. If we assume that the estimate is normally distributed then we can derive the interval by integrating Equation (14.5) between the limits of the range.

We often define the confidence interval in terms of a multiple (n) of the standard deviation. Figure 14.7 illustrates the case when n is 1. The probability of lying within one standard deviation of the mean is given by the shaded area under the probability density function. Figure 14.8 is the plot of *cumulative probability*, derived by applying the trapezium rule to determine the area under the probability density function. The shaded area is thus determined as 0.683 and the confidence interval in this example is therefore 68.3%. In general, the probability (P) of the value (x) being between $\mu - n\sigma$ and $\mu + n\sigma$ is given by

$$P = \int_{\mu - n\sigma}^{\mu + n\sigma} y.dx \tag{14.27}$$

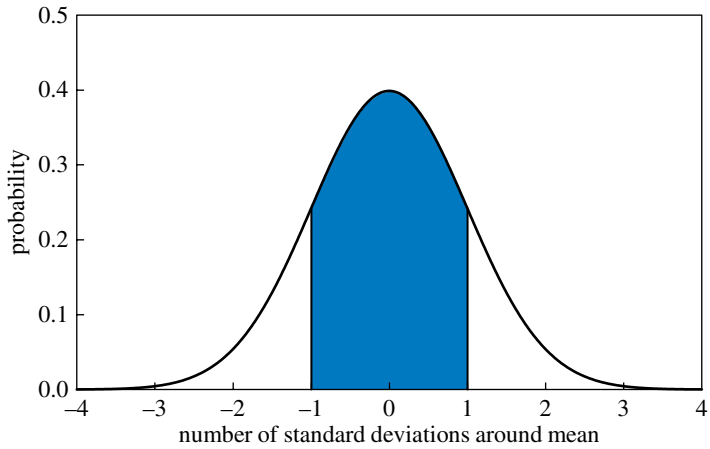

Figure 14.7 *Confidence interval on probability density plot*

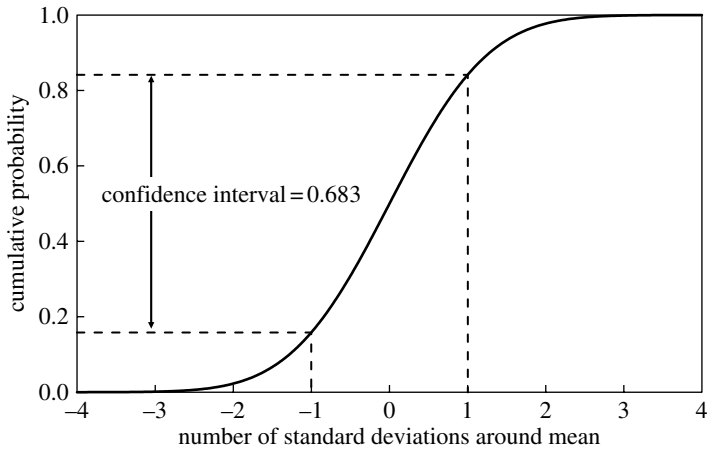

Figure 14.8 *Confidence interval on cumulative probability plot*

Since the distribution is symmetrical we can re-write Equation (14.27) as

$$P = 2\int_{\mu}^{\mu+n\sigma} y.dx \tag{14.28}$$

To perform the integration we first make, into Equation (14.5), the substitution

$$x = \mu + \sigma\sqrt{2}u \tag{14.29}$$

Therefore

$$P = \frac{2}{\sqrt{\pi}}\int_{0}^{n/\sqrt{2}} e^{-u^2}.du \tag{14.30}$$

Although it has nothing to do with errors, this result is known as the *error function* (*erf*). It is possible, although complicated, to integrate it but an easier approach is to use published look-up tables or the function provided in most spreadsheet packages.

$$P = \text{erf}\left(\frac{n}{\sqrt{2}}\right) \tag{14.31}$$

This function is plotted as curve A in Figure 14.9. So, for example, the 95% confidence interval is often quoted as 2σ (strictly 1.96σ). Table 14.1 shows some commonly used values. It should be remembered that these have been derived from the function for the normal distribution. There are other methods of determining the confidence interval in situations where the distribution is not normal. For any unimodal distribution the *Vysochanskij–Petunin inequality* states that the probability of a result lying more than $n\sigma$ from the mean will be less than $4(3n)^{-2}$. This is plotted as curve B in Figure 14.9 – showing the 95% confidence interval increases to 3σ, whereas 2σ now corresponds to a confidence interval of 89%. If not unimodal the *Chebyshev's inequality*, which applies to any distribution, states that the probability will be less than n^{-2}. This is plotted as curve C in Figure 14.9 – showing a further increase to 4.5σ, whereas 2σ now corresponds

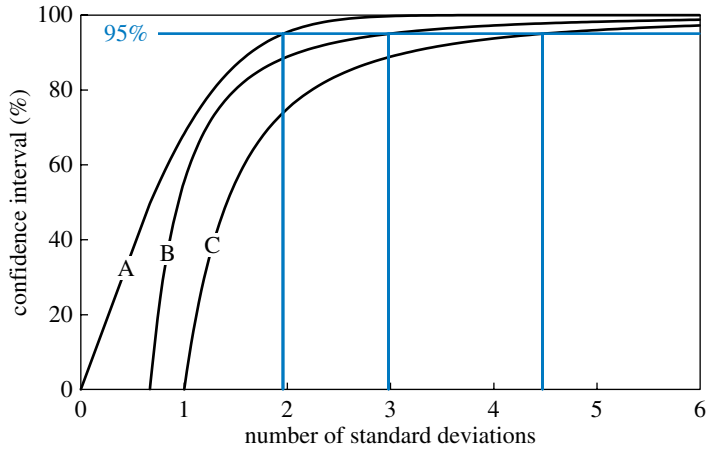

Figure 14.9 *Confidence interval as a function of the number of standard deviations*

Table 14.1 *Commonly used confidence intervals*

n	P (%)	P (%)	n
0.0	0	50	0.6745
0.5	38.3	60	0.8416
1.0	68.3	70	1.0364
1.5	86.6	80	1.2816
2.0	95.4	85	1.4395
2.5	98.76	90	1.6449
3.0	99.73	95	1.9600
3.5	99.954	98	2.3263
4.0	99.9937	99	2.5758
4.5	99.99932	99.5	2.8070
5.0	99.999943	99.9	3.2905
5.5	99.9999962	99.99	3.8906
6.0	99.9999998	99.999	4.4172
6.5	99.999999992	99.9999	4.8916
7.0	99.9999999997	99.99999	5.3267

to a confidence of 75%. As we shall see in the next few sections, there are other distributions which will generate different results. In certain circumstances, these may be more applicable.

14.8 Westinghouse Electric Company Rules

We discuss later in this chapter techniques for identifying data that might be considered as *outliers* and so be rejected from statistical analysis. The control engineer might apply such a method to exclude suspect data from the development of an inferential. A similar approach can be taken in identifying suspect results from an inferential once in operation. The development of the inferential will have quantified the expected σ_{error}. If, for example, the error between the inferential and the laboratory result exceeds three times this value then from Table 14.1 we can determine that the probability that the inferential is incorrect is 99.73%. The probability that it is correct is therefore very low at 0.27%. This would prompt an investigation of the reliability of the inferential and a possible update. This is the principle behind the *Shewhart statistical process control (SPC) chart*. This includes a centre line, drawn at the mean (where the mean would be 0 if the chart is being used to assess the error in an inferential) and the lower and upper *control limits* usually drawn at $\pm 3\sigma_{error}$. The inferential error is then plotted against sample number.

This simple approach is unlikely, however, to detect small shifts in the accuracy. A more sophisticated approach is based on the *Westinghouse Electric Company (WECO) rules* that were originally developed to

improve the reliability of Shewhart charts. If any of these four rules are violated then the inferential would be considered suspect.

1. The error exceeds the $3\sigma_{error}$ limit – as described above.
2. Two out of three consecutive errors exceed the $2\sigma_{error}$ limit but are within the $3\sigma_{error}$ limit. The errors exceeding the limit must do so on the same side of the mean. When the inferential was developed the mean error would have been 0 and so we need either consecutive positive errors or consecutive negative errors. From Table 14.1, the probability of being within the lower limit is 95.4%. One of the three errors must be in this region. The probability of being between the limits on the same side of the mean is given by (99.73 – 95.4)/2 or 2.17%. Two of the errors must be in this region. There are three combinations of results that meet the criteria and each can occur on either side of the mean. The probability of violating this rule is therefore given by $0.954 \times 0.0217^2 \times 3 \times 2$ or 0.27%.
3. Four out of five consecutive errors (with same sign) exceed the σ_{error} limit but are within the $2\sigma_{error}$ limit. Taking the same approach, the probability of violating this rule is given by $0.683 \times 0.1355^4 \times 5 \times 2$ or 0.23%.
4. Eight consecutive points fall on the same side of the mean. The probability of violation of this rule is 0.5^8 or 0.39%.

Applying only the first rule would result in a false indication of a problem once in every 371 instances. While increasing the reliability of detecting a problem, applying all four rules will increase the false indication rate to once in every 92 instances.

These rules all assume that the prediction error is normally distributed, so it would be wise to confirm this before applying them. For example, it is possible to define asymmetric rules for skewed distributions.

14.9 Gamma Function

The *gamma function* (Γ) is a way of more simply presenting the lengthy multiplication of a series of values that arise in defining probability distributions. A complete definition of the function is beyond the scope of this book but it is determined by applying the formula

$$\Gamma(x) = (x-1)\Gamma(x-1) = (x-1)(x-2)\Gamma(x-2) = \ldots \tag{14.32}$$

The series terminates when the last term $(x - n)$ is less than or equal to 1. So, if x is an integer, the last term in the multiplication will be $\Gamma(1)$, which is defined as 1. For example,

$$\Gamma(4) = (4-1)(3-1)(2-1)\Gamma(1) = 6 \tag{14.33}$$

For integers, $\Gamma(x)$ is simply the factorial of $(x - 1)$ – shown as points on the curve in Figure 14.10. But x may not be an integer; for example, if x is an odd multiple of 0.5 then the last term in the multiplication will be $\Gamma(0.5)$ – which is defined as $\pi^{0.5}$. For example,

$$\Gamma(3.5) = (3.5-1)(2.5-1)(1.5-1)\Gamma(0.5) = 1.875\pi^{0.5} \tag{14.34}$$

Fortunately the formulae for most distributions do not involve the gamma function of other forms of non-integer value. However, if needed, most spreadsheet packages include the function that can be used for any positive value of x.

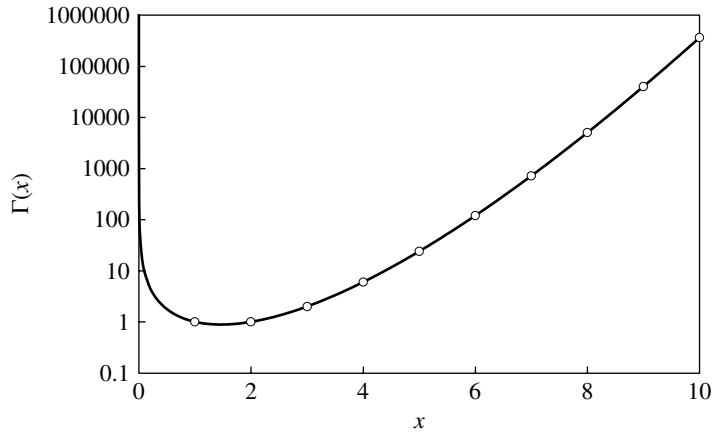

Figure 14.10 *Gamma function*

14.10 Student *t* Distribution

The *Student t distribution* was proposed by Gosset using the pseudonym Student. It is used when the mean of the population is estimated from a very small number of results. The method gives the reliability of the estimate. This is useful, for example, in determining whether a few unusual results are representative of the normal behaviour of the process or whether they indicate a problem.

As we showed previously in defining the calculation of standard deviation (Section 14.4), if σ is the standard deviation of a sample of n values then the mean ($\bar{x}$) of the sample will have a standard deviation of $\sigma / \sqrt{n}$. If μ is the mean of the population then the Student t is defined as

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \tag{14.35}$$

The probability density function $f(t)$ is defined as

$$f(t) = \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)\sqrt{\pi f}}\left(1 + \frac{t^2}{f}\right)^{-\frac{f+1}{2}} \tag{14.36}$$

To apply this function we first need to determine the *number of degrees of freedom* (f). This is in effect the number of values in our sample which are free to vary. It is determined by subtracting, from the total number of values (n), how many are fixed by any intermediate steps in the calculation. For example, we use the n values to calculate the mean. We can vary $n-1$ of the values as we wish provided the remaining value is adjusted so that the mean remains unchanged. The number of degrees of freedom is therefore $n-1$. The shape of the distribution depends on the number of degrees of freedom and hence on the sample size.

If f is an odd number then

$$\frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)\sqrt{\pi f}} = \frac{(f-1)(f-3)....\times 2}{\pi\sqrt{f}\,(f-2)(f-4)....\times 3} \tag{14.37}$$

If f is an even number then

$$\frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)\sqrt{\pi f}} = \frac{(f-1)(f-3)....\times 3}{2\sqrt{f}\,(f-2)(f-4)....\times 2} \tag{14.38}$$

If f is 1, Equation (14.36) gives

$$f(t) = \frac{1}{\pi\left(1+t^2\right)} \tag{14.39}$$

If f is 2, Equation (14.36) gives

$$f(t) = \frac{1}{\left(2+t^2\right)^{1.5}} \tag{14.40}$$

Figure 14.11 shows how the distribution for $f = 1$ compares to the normal distribution with a variance of 1. As f increases, the t distribution becomes closer to the normal distribution. The two distributions are identical for an infinite number of degrees of freedom. For values of f greater than 15 the distribution will be close enough to the normal distribution to make this approximation.

For example, imagine we are making a product which has an inferential property that, over the last 24 hours, shows that the property has been well-controlled at 100 and so μ is 100. In the same period three eight-hourly laboratory tests gave the results 98, 100 and 96. The mean of these results ($\bar{x}$) is 98 and the standard deviation (σ), from Equation (14.18), is 2. From Equation (14.35) t is given by

$$t = \frac{98-100}{2/\sqrt{3}} = -1.732 \tag{14.41}$$

We have calculated one mean and so two degrees of freedom remain. Figure 14.12 shows the cumulative t distribution for $f = 2$ – drawn by applying Equations (14.36) and (14.40). It shows that, if the inferential

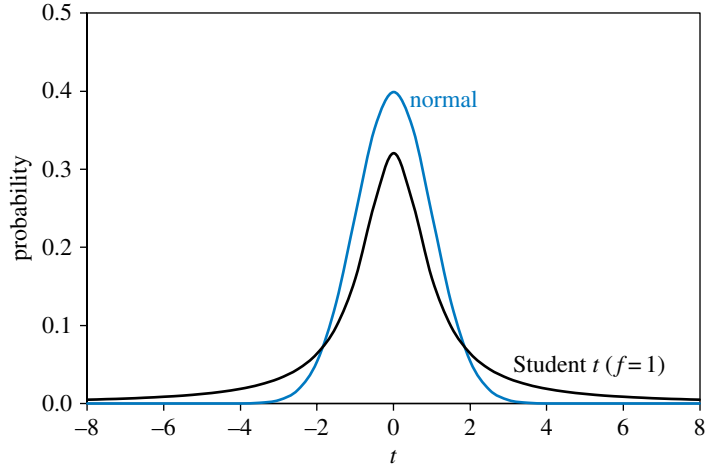

Figure 14.11 *Probability density for Student t distribution*

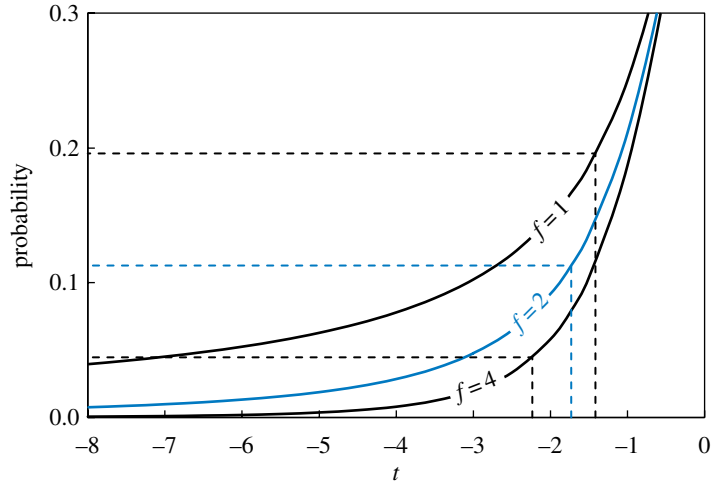

Figure 14.12 *Cumulative probability for Student t distribution*

is accurate, there is a probability of 0.113 that the true mean could be less than 98. One could thus conclude that this is within the 95% confidence that the inferential is correct. Figure 14.12 shows, had there been only two samples ($f = 1$) with the same mean and standard deviation, using Equations (14.36) and (14.39), the probability that the mean could be less than 98 rises to 0.196. Similarly, had there been five samples ($f = 4$), using Equations (14.36) and (14.38) shows it would fall to less than 0.05 – the level of probability that would typically prompt a redesign of the inferential. The *t* distribution therefore reflects our natural inclination to place more trust in a conclusion that is based on more results.

Another application of the *t* distribution is comparing the means of two data sets to determine whether they are significantly different. One set has n_1 measurements, a mean of $\bar{x}_1$ and a standard deviation of σ_1. The other has n_2 measurements, a mean of $\bar{x}_2$ and a standard deviation of σ_2. The standard deviation of the difference between the two means (σ_d) is given by

$$\sigma_d^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \tag{14.42}$$

And so *t* can be derived using Equation (14.35).

$$t = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma_d} \right| \tag{14.43}$$

On the basis that we have derived two values (the two means), the number of degrees of freedom (f) is given by

$$f = n_1 + n_2 - 2 \tag{14.44}$$

This might be used to determine, from a small number of samples, whether an inferential has developed a significant bias error. One set of data will be laboratory results and the other the corresponding estimates made by the inferential.

14.11 χ^2 Distribution

The χ^2 distribution is useful when we wish to compare a set of observed results with those predicted and to determine the probability that our method of predicting the results is correct. Like the t distribution, it uses the number of degrees of freedom (f). The probability density function $f(x)$ is defined as

$$f(x) = \frac{x^{0.5f-1}e^{-0.5x}}{2^{0.5f}\Gamma\left(\frac{f}{2}\right)} \tag{14.45}$$

If f is odd then

$$2^{0.5f}\Gamma\left(\frac{f}{2}\right) = \sqrt{2\pi}\,(f-2)(f-4)....\times 3 \tag{14.46}$$

If f is even then

$$2^{0.5f}\Gamma\left(\frac{f}{2}\right) = 2(f-2)(f-4)....\times 2 \tag{14.47}$$

Figures 14.13 and 14.14 show this function for a range of values of f. To use the distribution, we first calculate χ^2 from the predicted values ($\hat{y}$) and the observed values (y).

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} \tag{14.48}$$

Figure 14.15 shows the cumulative probability curves, plotted by applying the trapezium rule to Equation (14.45), for a range of values of f. From the appropriate curve we identify the probability corresponding to the value of χ^2.

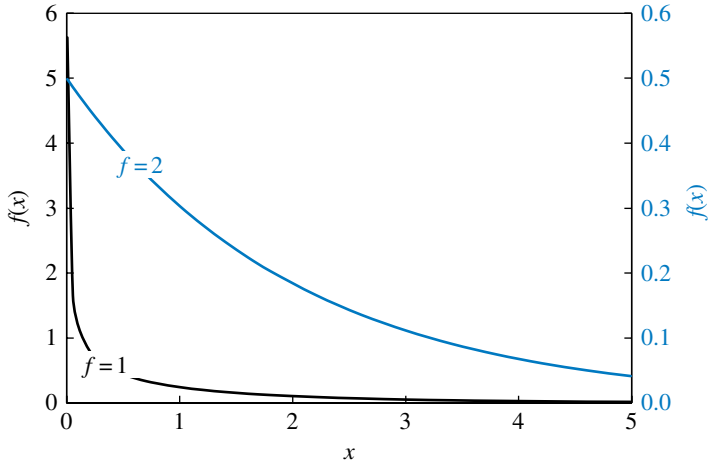

Figure 14.13 *χ^2 distribution for f = 1 and f = 2*

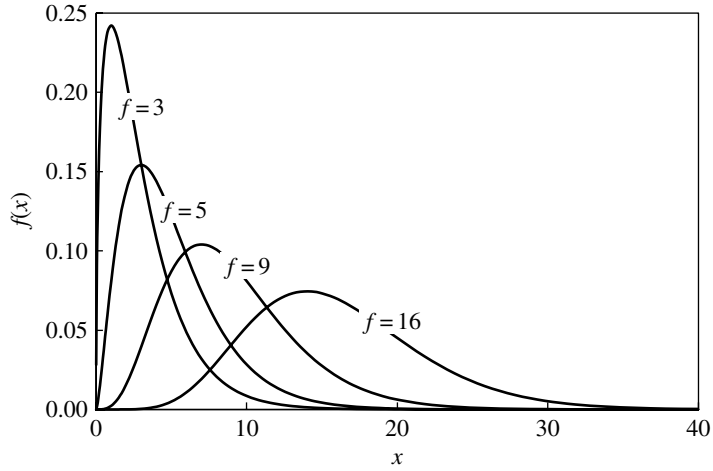

Figure 14.14 *χ² distribution for f > 2*

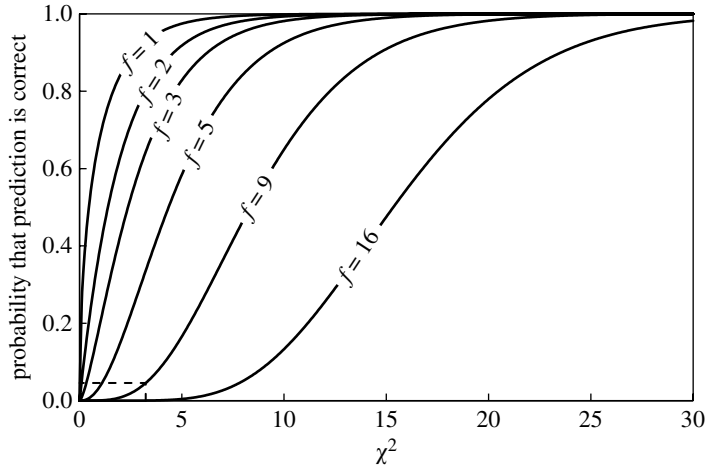

Figure 14.15 *Cumulative probability for χ² distribution*

As an example we will use a correlation developed using regression analysis shown later in this chapter. The control engineer might think of this as an inferential property calculation.

$$\hat{y} = 25.76 + 1.468x \tag{14.49}$$

Table 14.2 shows daily measurements of the true property (y) versus those predicted by the inferential. The same information is portrayed in Figure 14.16. We can see that every measurement is greater than that predicted and so we wish to determine whether the inferential calculation should be updated. If we consider all 10 data points then f is 9 and χ^2 is 3.24. We can see from Figure 14.15 that the probability that our inferential property calculation is correct is 0.046. This means we are 95.4% sure that the inferential has

Table 14.2 *Measured and predicted values for y*

x	y	$\hat{y}$	$\frac{(y-\hat{y})^2}{\hat{y}}$	$\sum\frac{(y-\hat{y})^2}{\hat{y}}$	f	P
11	45	42	0.23	0.23	0	
13	49	45	0.39	0.61	1	0.566
17	56	51	0.55	1.16	2	0.441
27	71	65	0.48	1.64	3	0.351
35	80	77	0.11	1.75	4	0.218
49	105	98	0.55	2.30	5	0.193
19	56	54	0.10	2.40	6	0.120
52	109	102	0.47	2.87	7	0.103
65	127	121	0.28	3.15	8	0.075
66	126	123	0.09	3.24	9	0.046

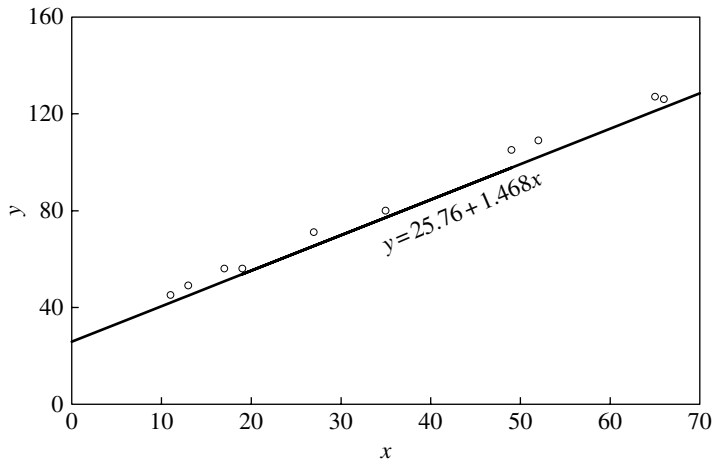

Figure 14.16 *Assessing reliability of predicted values*

been incorrect and so we would work to resolve the problem. Had we performed the analysis a day sooner, when f was 8 and χ^2 was 3.15, the probability that the inferential was incorrect was 92.5% – just below the level where we would investigate.

Following the fourth Westinghouse rule, described in Section 14.8, we would have taken action two days sooner – once eight consecutive errors had the same sign. But this criterion takes no account of the size of the error and so is perhaps too crude.

14.12 *F* Distribution

Named in honour of Fisher, the *F distribution* is the distribution of the ratio of the estimates of the variance of a normal distribution. While at first this might seem a purely theoretical function it has, as we shall see later, a range of applications in the *analysis of variance* (*ANOVA*). In particular we can use it to determine whether a measured change in variance is significant.

The mathematical function that describes its shape is quite complex, involving two values for the number of degrees of freedom – f_1 for the numerator and f_2 for the denominator of the ratio. Fortunately, as with the distributions already covered in the chapter, it has been converted by others into tables and spreadsheet functions.

$$f(x) = \frac{\Gamma\left(\frac{f_1 + f_2}{2}\right)}{\Gamma\left(\frac{f_1}{2}\right)\Gamma\left(\frac{f_2}{2}\right)}\left(\frac{f_1}{f_2}\right)^{0.5f_1} x^{0.5f_1 - 1}\left(1 + \frac{f_1}{f_2}x\right)^{-0.5(f_1 + f_2)} \tag{14.50}$$

If f is odd then, from Equation (14.46)

$$\Gamma\left(\frac{f}{2}\right) = \frac{\sqrt{2\pi}\,(f-2)(f-4)....\times 3}{2^{0.5f}} \tag{14.51}$$

If f is even then, from Equation (14.47)

$$\Gamma\left(\frac{f}{2}\right) = \frac{2(f-2)(f-4)....\times 2}{2^{0.5f}} \tag{14.52}$$

To show the function graphically is also more complex since it depends on the values of two parameters (f_1 and f_2). Figures 14.17 to 14.19 give some indication of how the shape of the curve varies.

A useful application of the *F* distribution is as part of multivariable regression analysis. For example, when developing inferential properties, we have to decide which independent variables should be included. Adding a variable will always improve the apparent accuracy of the inferential but we should assess whether the improvement makes the additional complexity worthwhile. For example, Table 14.3 shows a series of measurements collected for a C_3/C_4 splitter. The aim is to build an inferential for the $\%C_4$ in the overhead propane product (y). In this example we have measurements (x_1 and x_2) of the temperature of two trays close to the top of the 20-tray column – on trays 15 and 17. As a measure of the accuracy of the predicted value ($\hat{y}$) we will use the variance of the error (σ^2_{error}), defined as

$$\sigma^2_{error} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n} \tag{14.53}$$

We will see in a later section how the regression is performed. In this example, depending on which temperature is used, we obtain two possible inferentials.

$$\hat{y} = -56.38 + 0.9253x_1 \tag{14.54}$$

$$\hat{y} = -28.09 + 0.5494x_2 \tag{14.55}$$

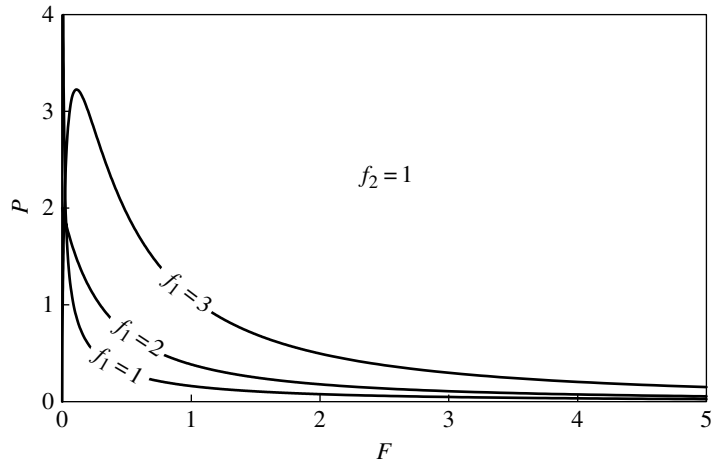

Figure 14.17 *F function for $f_2 = 1$*

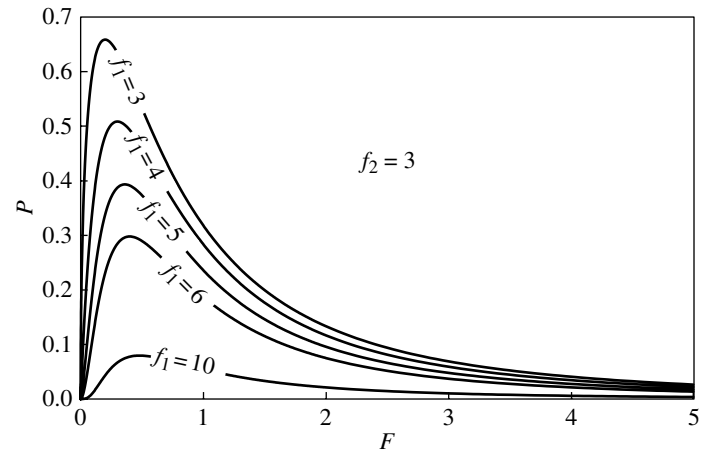

Figure 14.18 *F function for $f_2 = 3$*

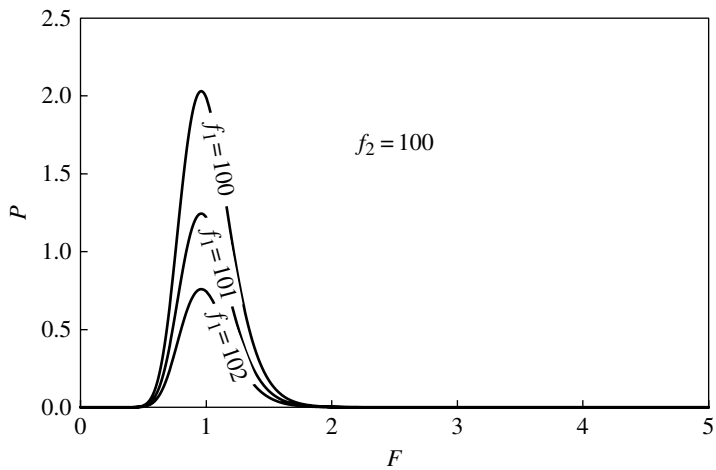

Figure 14.19 *F function for $f_2 = 100$*

Table 14.3 *Selection of one or two parameter model*

Tray 15 (x_1)	Tray 17 (x_2)	$\%C_4$ (y)
66.5	60.8	5.34
64.9	58.0	3.79
66.4	59.4	4.28
67.4	62.5	6.46
66.2	57.6	3.41
66.3	60.5	5.26
67.2	61.9	5.91
67.2	62.1	5.86
66.8	61.1	5.43
65.1	58.2	3.90
65.6	59.0	4.41
65.4	58.7	4.13
66.2	60.1	4.80
64.3	57.0	3.48
66.4	60.4	5.08
67.7	62.8	6.61
66.9	61.5	5.70
66.4	60.5	5.08
65.9	59.5	4.37
64.6	57.5	3.67

The first of these has a σ^2_{error} of 0.1775, the second 0.0198. Of these we would therefore clearly choose the second. The performance of both is shown in Figure 14.20. As expected the temperature of the tray nearer the top of the column gives the better prediction. However, we could use both temperatures in the inferential.

$$\hat{y} = -20.54 - 0.2092x_1 + 0.6544x_2 \tag{14.56}$$

This has a σ^2_{error} of 0.0150. As expected it is better than either of the single-input possibilities. Its performance is shown in Figure 14.21. However, we would expect x_1 to be highly correlated with x_2 and so add little value. Indeed, as Figure 14.22 shows, this is the case for most of the measurements with the exception of two – which might be considered outliers. This explains the relatively small improvement in σ^2_{error}.

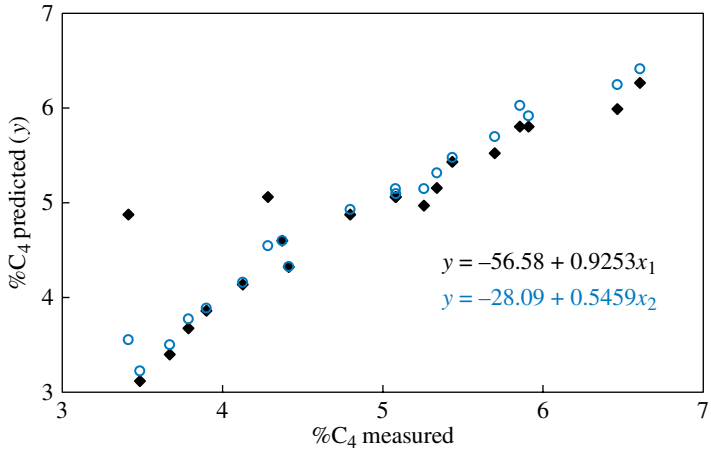

Figure 14.20 *Single-input predictions*

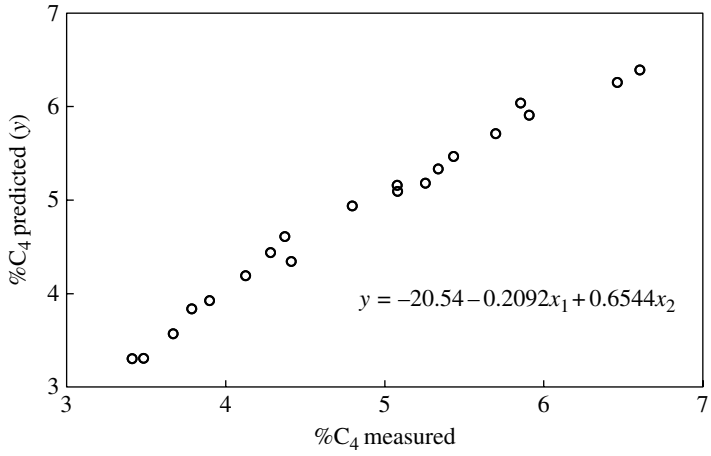

Figure 14.21 *Two-input prediction*

We need to check whether including the second temperature gives a significantly better prediction. To do so we first calculate F which, by definition, is given by

$$F = \frac{\left[\dfrac{\sigma_1^2 - \sigma_2^2}{f_1}\right]}{\left[\dfrac{\sigma_2^2}{f_2}\right]} \tag{14.57}$$

In our example σ_1^2 is that for the better of the two single-input inferentials and σ_2^2 is that for the two-input version. σ_1^2 will always be greater than σ_2^2. The degrees of freedom are derived from the number of

coefficients (p) used in the inferential. In this example p_1 is 2 and p_2 is 3. The total number of measurements (n) in our example is 20.

$$f_1 = p_2 - p_1 \tag{14.58}$$

$$f_2 = n - p_2 \tag{14.59}$$

$$\therefore \quad F = \frac{\left[\dfrac{0.0198 - 0.0150}{3 - 2}\right]}{\left[\dfrac{0.0150}{20 - 3}\right]} = 5.44 \tag{14.60}$$

Figure 14.23 shows a selection of cumulative probability curves for different combinations of f_1 and f_2. We make the null hypothesis that the two-input correlation is not significantly different from the

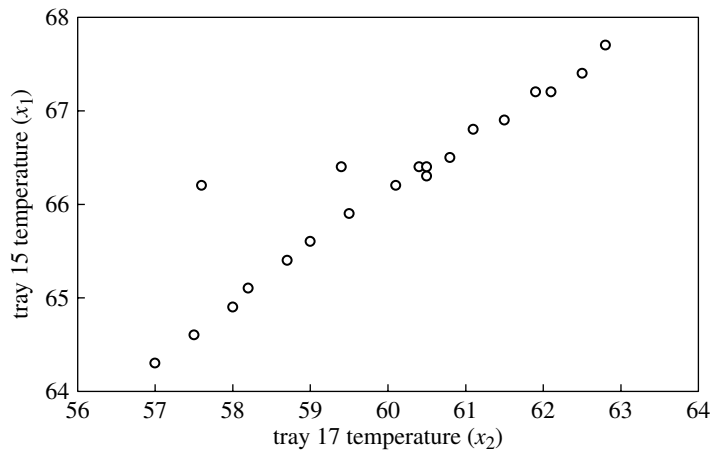

Figure 14.22 *Correlation between inputs*

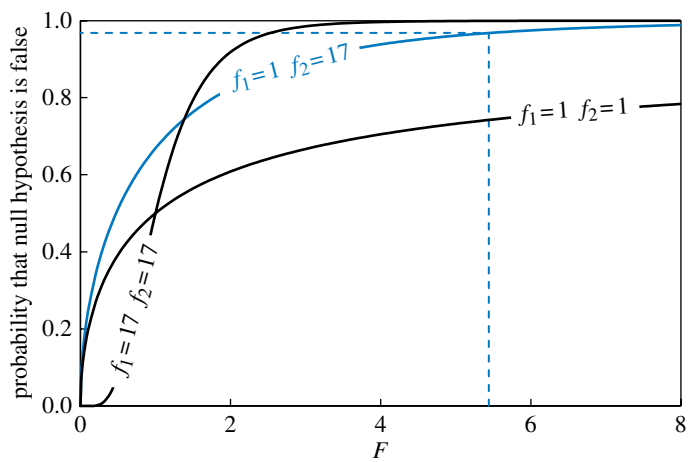

Figure 14.23 *Cumulative probability for F distribution*

single-input alternative. The curves have been plotted to give the probability that the null hypothesis is false. In our example they show a probability of around 95% that this is the case – in other words we should adopt the two-input inferential.

The alert engineer might at first be suspicious of the end result. We would expect $\%C_4$ to increase as the tray temperatures increase and so would consider suspect the negative coefficient for x_1 in Equation (14.56). But, as covered in Chapter 12, we can rewrite this equation.

$$\hat{y} = -20.54 + 0.4452x_2 - 0.2092\left(x_1 - x_2\right) \tag{14.61}$$

This is consistent with Equation (14.55) which shows that the tray 17 temperature (x_2) is the better choice. Since x_1 is a temperature lower down the column it will be greater than x_2 and so ($x_1 - x_2$) will always be positive. This difference is a measure of the separation taking place in this section of the column. Increasing separation will reduce $\%C_4$ and so the negative coefficient for this parameter is correct.

14.13 Akaike Information Criterion

The *Akaike Information Criterion* (*AIC*) offers an alternative method of deciding whether the inclusion of additional parameters in a correlation gives a statistically significant improvement in its accuracy. There are several published definitions but, if the correlation is developed using the conventional least squares regression and the prediction error is normally distributed, then

$$AIC = n\ln\left(\sigma^2_{error}\right) + 2p \tag{14.62}$$

As usual n is the number of data points; p is the number of independent variables used in the correlation. If n is small compared to p it is advisable to use the *second order AIC* to avoid over-fitting.

$$AIC = n\ln\left(\sigma^2_{error}\right) + \frac{2pn}{n-p-1} \tag{14.63}$$

Applying this technique to the data used in the previous section, the correlation described by Equation (14.55) has an *AIC* of −76.2, given by

$$AIC_1 = 20\ln\left(0.0198\right) + \frac{20\times 2\times 1}{20-1-1} \tag{14.64}$$

That described by Equation (14.56) has an *AIC* of -79.3.

$$AIC_2 = 20\ln\left(0.0150\right) + \frac{20\times 2\times 2}{20-2-1} \tag{14.65}$$

The better correlation is the one with the lower *AIC*. Whether the difference is significant is determined from ΔAIC, given by

$$\Delta AIC = AIC_1 - AIC_2 = 3.1 \tag{14.66}$$

A value less than 2 for ΔAIC indicates that increasing the number of independent variables is probably not justified by the improvement in accuracy. The larger value of 3.1 suggests that the additional variable

is valuable, agreeing with the conclusion of the F test in the previous section. Very large values, those greater than 10, indicate that there are variables that are better excluded from the original correlation.

An alternative comparison is to determine the relative probability, given by

$$\exp\left[\frac{-\Delta AIC}{2}\right] = 0.22 \tag{14.67}$$

This indicates that there is a probability of 22% that the inclusion of the additional parameter will minimise the loss of information. Alternatively we are only 78% certain that the inferential is not improved by the addition of the second variable. On this basis we would also elect to include it.

If n is significantly larger than p then we can apply Equation (14.62) to determine the condition necessary to justify the inclusion of one additional parameter. We require

$$AIC_p - AIC_{p+1} > 2 \tag{14.68}$$

Therefore

$$\left(n\ln\left(\sigma_p^2\right)+2p\right)-\left(n\ln\left(\sigma_{p+1}^2\right)+2\left(p+1\right)\right)>2 \tag{14.69}$$

Rearranging

$$\ln\left(\frac{\sigma_p^2}{\sigma_{p+1}^2}\right)-\frac{2}{n}>0 \qquad \text{or} \qquad \ln\left(\frac{\sigma_p}{\sigma_{p+1}}\right)-\frac{1}{n}>0 \tag{14.70}$$

In our example the left hand side of the second inequality evaluates to 0.178, again confirming that including the additional parameter is justified. If required, a similar but more complex inequality can be derived from Equation (14.63).

While normally the number of inputs to an inferential would be increased incrementally with the AIC test applied at each stage, this may not always be possible. For example, we might wish to compare an inferential that uses three inputs with one that uses only one (which is not one of the three). Equation (14.70) can be extended to handle this. For an inferential using p_1 inputs to be preferable to one using p_2 inputs,

$$\ln\left(\frac{\sigma_{p_1}}{\sigma_{p_2}}\right)-\frac{p_2-p_1}{n}>0 \tag{14.71}$$

The technique can also be applied to dynamic model identification. As described later in this chapter, model order is determined by the number of historical values of PV and MV included. Increasing the order will always increase the accuracy of the model. Using the AIC test as the order is incremented will ensure that the order is not increased to the level where only noise is being modelled.

The technique can also be used to determine whether a correlation has become more reliable by developing it from a larger number of data points. If the original was developed from n points and the revised version developed from m points then, for it to be significantly better

$$\ln\left(\frac{\sigma_n}{\sigma_m}\right)-\frac{m}{2n}>0 \tag{14.72}$$

14.14 Adjusted R^2

From Equation (14.25), Pearson R^2 is defined by

$$R^2 = \frac{\left(\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i}^{n}(y_i - \bar{y})^2} \tag{14.73}$$

Since the addition of an input will always improve a correlation the normal Pearson R^2 cannot be used, if it uses a different number of inputs, to properly compare correlations. Instead we should use the *adjusted* R^2 – usually written as $\bar{R}^2$. The principle behind the adjusted version is to include a penalty as additional inputs are used, where n is the number of data points and p the number of inputs.

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1} \tag{14.74}$$

If n is 2 we have two data points (x_1, y_1) and (x_2, y_2), and so

$$\bar{x} = \frac{x_1 + x_2}{2} \qquad \text{and} \qquad \bar{y} = \frac{y_1 + y_2}{2} \tag{14.75}$$

Equation (14.73) will then give a result of 1 for R^2. If there is only one input ($p = 1$) then $\bar{R}^2$ will be indeterminate – although usually then assumed to be zero. In general n must be greater than $p + 1$.

Unlike R^2, $\bar{R}^2$ can be less than 0. For example, if we retain only the one input (but this time with 3 data points) then, if R^2 is at its lowest value of zero, from Equation (14.74), $\bar{R}^2$ will be at its lowest possible value of -1. In general, if R^2 is 0 and n is at its lowest possible value of $p + 2$, $\bar{R}^2$ will be $-p$.

An approach similar to Akaike can be derived. If we increase the number of inputs from p to p+1 then for the correlation to have genuinely improved

$$1 - \left(1 - R_{p+1}^2\right)\frac{n-1}{n-(p+1)-1} > 1 - \left(1 - R_p^2\right)\frac{n-1}{n-p-1} \tag{14.76}$$

Or

$$R_{p+1}^2 > \frac{1 + (n-p-2)R_p^2}{n-p-1} \tag{14.77}$$

Using the same correlations compared by Akaike, the left hand side evaluates to 0.991 while the right is 0.979, showing again that the improvement justifies the use of an additional input. As with Akaike, adjusted R^2 can also be used to determine the best choice of order for dynamic models.

The adjusted R^2 can similarly be used to determine whether using a larger number of data points to develop the correlation has produced a significantly more reliable version. If the original was developed from n points and the revised version developed from m points then, for it to be significantly better

$$R_m^2 > \frac{(m-p-1)(n-1)R_n^2 - p(m-n)}{(n-p-1)(m-1)} \tag{14.78}$$

14.15 Levene's Test

Levene's test offers another approach to determining the significance of any difference in the variances of multiple sets of data. We have a total of n values split between k sets of data, where n_i is the number of values in the i^{th} set, and so

$$\sum_{i=1}^{k} n_i = n \tag{14.79}$$

For every value (y) we determine a corresponding value of z. For the j^{th} value of the i^{th} set, where $\tilde{y}_i$ is the median of the i^{th} set,

$$z_{ij} = \left| y_{ij} - \tilde{y}_i \right| \tag{14.80}$$

Levene originally used the mean ($\overline{y}_i$), which works well if the data are normally distributed. The use of the median (the 50% percentile) was proposed by Brown and Forsythe [42] as more robust because it is much less affected by outliers. Other approaches include the use of the *trimmed mean* where only values between the 5% and 95% percentiles are used in its calculation.

We then calculate the mean ($\overline{z}$) of all the z_{ij}. We also determine the means ($\overline{z}_i$) of each of the i sets.

$$\overline{z} = \frac{1}{n} \sum_{i=1}^{k} \left[\sum_{j=1}^{n_i} z_{ij} \right] \tag{14.81}$$

$$\overline{z}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} z_{ij} \tag{14.82}$$

The results are then used to calculate the value of W.

$$W = \frac{n-k}{k-1} \frac{\sum_{i=1}^{k} n_i \left(\overline{z}_i - \overline{z} \right)^2}{\sum_{i=1}^{k} \left[\sum_{j=1}^{n_i} \left(z_{ij} - \overline{z}_i \right)^2 \right]} \tag{14.83}$$

We make the null hypothesis that the variances of the populations are equal. We then apply the F test against a probability (typically 0.95) with degrees of freedom f_1 and f_2 given by

$$f_1 = k - 1 \tag{14.84}$$

$$f_2 = n - k \tag{14.85}$$

If W exceeds F we reject the hypothesis and conclude that the variances of the sets of data are significantly different.

To illustrate the method, we will again evaluate the inferentials given by Equations (14.55) and (14.56). Since we are comparing two sets of data k is 2. Both n_1 and n_2 are 20, and n is therefore 40. We cannot simply compare the variance of the predictions made by each of the methods. To understand this, imagine that we have two inferentials which give exactly the same result but we shuffle the sequence of the results

of one of them. Shuffling does not affect its variance (or mean or median) and so the analysis would conclude that the accuracy of the inferential is unaffected by the sequence in which it generates predictions. Instead we must compare the variances of the prediction error (y) for each inferential.

The values $y_{1,1}$ to $y_{1,20}$ are the prediction errors arising from applying Equation (14.55) to the 20 sets of data. Similarly $y_{2,1}$ to $y_{2,20}$ are derived using Equation (14.56). The medians $\tilde{y}_1$ and $\tilde{y}_2$ can then be determined as −0.0033 and −0.0085 respectively. Using Equation (14.80), these values are used to derive z_{ij}. Equation (14.81) then gives 0.1029 for $\bar{z}$. Equation (14.82) gives 0.1095 for $\bar{z}_1$ and 0.0963 for $\bar{z}_2$. From Equation (14.83) we then obtain a value of 0.24 for W.

From Equation (14.84) we obtain a value of 1 for f_1 and, from Equation (14.85), a value of 38 for f_2. The cumulative probability curve for these degrees of freedom lies very close to the coloured line in Figure 14.23. The probability that the null hypothesis is false is 0.095 (or 9.5%). In other words there is a 90.5% probability that the improvement in σ^2_{error}, achieved by including the additional input, could easily be accounted for by chance. Levene is therefore in conflict with the methods applied in the previous two sections.

The performance parameter (ϕ) recommended by Equation (9.19) increases from 0.9773 to 0.9828. Using these values in Equation (9.21) gives

$$100\left[1-\sqrt{\frac{1-0.9828}{1-0.9773}}\right]=13 \tag{14.86}$$

The two-input inferential would permit the benefits captured by APC to be increased by about 13% compared to what would be achievable using the single-input version. This example shows that statistical methods should only be considered as a guide and are not a substitute for good engineering judgement. In this example, at least, the engineer should be guided by the adjusted R^2 and AIC, rather than by Levene.

14.16 Box-Wetz Ratio

The *Box-Wetz ratio* is used by some as an improvement on Pearson R^2. Like the performance index (ϕ) suggested in Chapter 9, its aim is to assess the variance of the prediction error in relation to the variation in the true value. It is given by

$$BW=\frac{\hat{y}_{max}-\hat{y}_{min}}{\sqrt{\frac{(p+1)s^2}{n}}} \tag{14.87}$$

The numerator represents the range of the predicted value. As usual p is the number of inputs used by the inferential and n the number of data points used in developing the inferential. The term s^2 is an estimate of the variance of the prediction error of the whole population, given by

$$s^2=\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-p-1} \tag{14.88}$$

The developers of the technique suggest that, for the correlation to be reliable, the value of BW should be at least 4. Others have suggested it should be greater than 10. Calculating it for the data in Table 9.1 gives values of 0.45 and 0.15 for the first two cases – suggesting that it is a reliable indicator of poor accuracy.

It gives a value of 0.30 for the third case but this requires only a single bias update to be perfectly accurate. Indeed, a sustained bias error of 0.38 would result in rejection at a value of *BW* of 4. Rejection at the level of 10 would result from a bias error of 0.15 – an error of around only 3%.

Figure 14.24 shows the effect, compared to Figure 9.11, of replacing ϕ with *BW*. Even against the minimum value of 10, *BW* would indicate that the correlation is always accurate. The large errors highlighted by the use of ϕ are virtually hidden by the noise in *BW*. So, while the Box-Wetz ratio may be an effective indicator for other applications, it would appear an unreliable method of assessing inferential properties. This, and the difficulty of linking the ratio to the economic benefit of improving an inferential, suggest it should not be used.

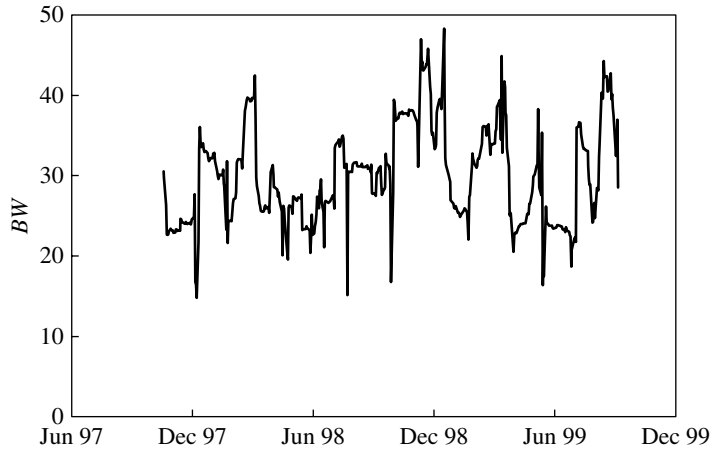

Figure 14.24 *Use of Box-Wertz ratio to assess inferential*

14.17 Regression Analysis

Figure 14.25 shows n data points (x, y). In this example we wish to identify the equation of the (straight) line of best fit, as shown, where a_0 is the intercept on the y axis and a_1 is the slope of the line, i.e.

$$\hat{y} = a_0 + a_1 x \tag{14.89}$$

The equation of the line is developed to minimise the residual sum of the squares (*RSS*) between the predicted value of y and the actual value, i.e.

$$\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2 = \sum_{i=1}^{n} \left(a_0 + a_1 x_i - y_i\right)^2 \tag{14.90}$$

Partially differentiating with respect to each of a_1 and a_0, and setting the derivative to 0 will identify the best choice of these values, i.e.

$$\frac{\partial \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{\partial a_0} = \sum_{i=1}^{n} 2\left(a_0 + a_1 x_i - y_i\right) = 0 \tag{14.91}$$

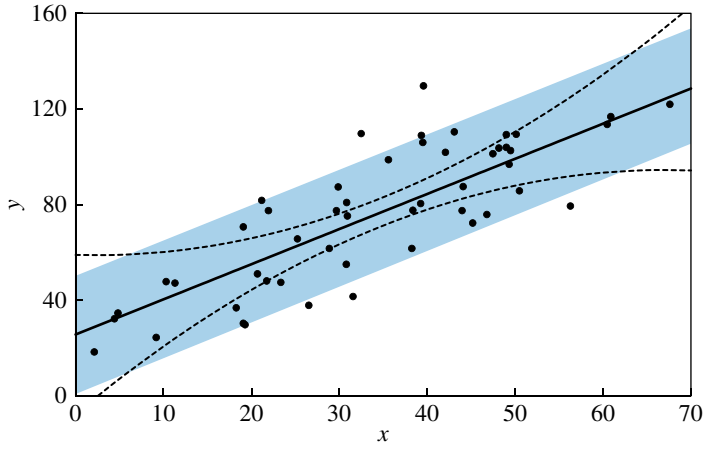

Figure 14.25 *Linear regression analysis*

$$\therefore \quad na_0 + a_1 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0 \qquad \text{or} \qquad a_0 + a_1 \bar{x} - \bar{y} = 0 \tag{14.92}$$

(Incidentally, calculating the mean of the predicted values from Equation (14.89) and comparing the result with Equation (14.92) shows that it is identical to the mean of the actual values.)

$$\frac{\partial \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{\partial a_1} = \sum_{i=1}^{n} 2x_i \left(a_0 + a_1 x_i - y_i\right) = 0 \tag{14.93}$$

$$\therefore \quad a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i = 0 \qquad \text{or} \qquad n a_0 \bar{x} + a_1 \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i = 0 \tag{14.94}$$

Solving Equations (14.92) and (14.94) gives

$$a_0 = \frac{\bar{y} \sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \qquad \text{and} \qquad a_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}.\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{14.95}$$

This line of best fit is plotted in Figure 14.25. The minimised *RSS* is given by

$$\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2 = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 - a_1^2 \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 \tag{14.96}$$

The last term of Equation (14.96) is the amount by which the variability of y has been explained by the variability of x.

This same approach is theoretically possible for more than one independent variable but a manual solution of the resulting equations would be impractical. Rewriting Equations (14.92) and (14.94) for the case of a single independent variable x_1

$$a_0 n + a_1 \sum x_1 = \sum y \tag{14.97}$$

$$a_0 \sum x_1 + a_1 \sum x_1^2 = \sum x_1 y \tag{14.98}$$

Applying the same approach we can derive similar equations for two independent variables, x_1 and x_2.

$$a_0 n + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \tag{14.99}$$

$$a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 = \sum x_1 y \tag{14.100}$$

$$a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 = \sum x_2 y \tag{14.101}$$

And for three independent variables, x_1, x_2 and x_3

$$a_0 n + a_1 \sum x_1 + a_2 \sum x_2 + a_3 \sum x_3 = \sum y \tag{14.102}$$

$$a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 + a_3 \sum x_1 x_3 = \sum x_1 y \tag{14.103}$$

$$a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 + a_3 \sum x_2 x_3 = \sum x_2 y \tag{14.104}$$

$$a_0 \sum x_3 + a_1 \sum x_1 x_3 + a_2 \sum x_2 x_3 + a_3 \sum x_3^2 = \sum x_3 y \tag{14.105}$$

We can also arrange these last four equations into matrix form.

$$\begin{pmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 & \sum x_1 x_3 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \sum x_2 x_3 \\ \sum x_3 & \sum x_1 x_3 & \sum x_2 x_3 & \sum x_3^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \\ \sum x_3 y \end{pmatrix} \tag{14.106}$$

The reader should now see the pattern by which higher numbers of independent variables would be incorporated. The coefficients (a) would then be determined by solving these equations.

The partial differentiation performed in developing the formulae above assumes that the variables are truly independent. If this is not the case then the resulting correlation may be suspect. Indeed, if there is a perfect correlation between two variables, the formulae will fail. Consider the case where there is a linear relationship between x_2 and x_1, such that

$$x_2 = p + q x_1 \tag{14.107}$$

Then, by substituting for x_2 in Equation (14.104), we obtain

$$a_0 \sum (p x_1 + q) + a_1 \sum x_1 (p x_1 + q) + a_2 \sum (p x_1 + q) x_2 + a_3 \sum (p x_1 + q) x_3$$

$$= \sum (p x_1 + q) y \tag{14.108}$$

This can be rewritten as

$$p\left(a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 + a_3\sum x_1x_3\right)$$
$$+q\left(a_0n + a_1\sum x_1 + a_2\sum x_2 + a_3\sum x_3\right)$$
$$= p\sum x_1y + q\sum y \tag{14.109}$$

But this could also have been derived by multiplying Equation (14.103) by p, multiplying Equation (14.102) by q and adding the results together. In other words we no longer have four independent equations with four unknown coefficients. In order to solve the equations, we have to set either a_1 or a_2 to zero – effectively removing either x_1 or x_2 from the analysis. One of these variables is *redundant* and the inclusion of both would result in one row of the Σ matrix in Equation (14.106) being derivable from others – thus making the matrix *singular*.

Because of random errors in both measurements, there will not be an exact correlation between two variables and a solution will exist. However, the values obtained for a_1 and a_2 will be unreliable and would likely show large variation if derived from different subsets of the data. It would be wise to remove one of the variables before progressing with the regression – retaining the one which is more reliably measured. It is good discipline in any case to first check for any cross-correlations between the independent variables to assess whether there is any redundancy. Process data, used to develop inferential properties, will usually show some level of cross-correlation. However, in practice, such weaker correlations tend not to undermine the value of the resulting inferential.

We can define a confidence interval for the resulting correlation. The standard deviation of the prediction error (σ) is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}-y\right)^2}{n-2}} \tag{14.110}$$

The denominator is $n-2$ rather than $n-1$ because we use two degrees of freedom to calculate intermediate values (a_0 and a_1). In the more conventional definition of standard deviation we only calculate one – the mean.

Without presenting the complex derivation, the confidence interval for the line of regression is given by

$$\sigma_y^2 = t.\sigma^2\left[\frac{1}{n} + \frac{\left(x-\bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\right] \tag{14.111}$$

Strictly t is from the Student t distribution, determined by the number of degrees of freedom and the required confidence interval. In our example we have 51 data points, giving 50 degrees of freedom. The 95% confidence interval would require a value of 2.0086 for t. The value approaches that of the normal distribution of 1.9600 shown in Table 14.1. Since regression analysis usually involves a large number of values, little accuracy would be lost by using this estimate. The confidence interval has been included in Figure 14.25 as the dashed lines. We might expect 95% of the points to lie within the dashed lines but only about 40% do so. This is because the lines are the boundaries of all possible straight lines, not the confidence interval of the predicted value. (To illustrate the difference, the 95% confidence interval for the

predicted value is shown as the shaded band.) Rapidly diverging boundaries are an indication that there is poor scatter and that points should be collected over a wider range.

These *confidence boundaries* for the line of regression can only be drawn based on a single independent variable. A correlation can of course be based on several independents. We can plot the boundaries for each independent by compensating its values to remove the effect of the other independents. Imagine that we have developed a correlation based on three independent variables.

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 \tag{14.112}$$

From Equation (14.92) we can see that the same correlation will apply to the means.

$$\bar{y} = a_0 + a_1 \bar{x}_1 + a_2 \bar{x}_2 + a_3 \bar{x}_3 \tag{14.113}$$

Subtracting gives:

$$y - \bar{y} = a_1 \left(x_1 - \bar{x}_1 \right) + a_2 \left(x_2 - \bar{x}_2 \right) + a_3 \left(x_3 - \bar{x}_3 \right) \tag{14.114}$$

Suppose we want to determine the confidence boundaries for the line of regression defined by a_2. Rearranging we get

$$y = a_2 \left[x_2 + \frac{a_1}{a_2} \left(x_1 - \bar{x}_1 \right) + \frac{a_3}{a_2} \left(x_3 - \bar{x}_3 \right) \right] + \bar{y} - a_2 \bar{x}_2 \tag{14.115}$$

We can therefore rewrite the correlation using the compensated independent variable $\left(x_2' \right)$.

$$x_2' = x_2 + \frac{a_1}{a_2} \left(x_1 - \bar{x}_1 \right) + \frac{a_3}{a_2} \left(x_3 - \bar{x}_3 \right) \tag{14.116}$$

We have adjusted each value of x_2 to reflect what they would need to be for the correlation to give the same prediction if the other independent values were fixed at their means. In fact we can choose any other reference value. This is applied to define linear pressure compensated temperatures used in the control of distillation columns, as described by Equations (9.8) and (12.67). Similarly, it is applied to develop weighted average bed temperatures that correlate with reactor conversion, as in Equation (9.12).

We can now write a correlation based on a single input and use this to determine the confidence interval for the line of regression.

$$y = a_0' + a_2 x_2' \qquad \text{where} \quad a_0' = \bar{y} - a_2 \bar{x}_2 \tag{14.117}$$

Further, if the plot of y against x_2' shows a nonlinear relationship, this can be used to help select a suitable nonlinear function that might be applied to x_2. Figure 14.26 shows an example where y is the concentration of heavy key component in the overhead product from a distillation column and x_2 is the temperature measured on a tray close to the top. The black points show that there appears to be no correlation. This is because of variations in x_1 – the column pressure. Applying Equation (14.116) removes the effect of this variation. The coloured points now show a clear correlation that suggests the PCT would be better included as a quadratic, rather than a linear, function.

It should be noted that regression does not tell us which is the independent variable and which is the dependent. It merely enables us to show that they are correlated. Indeed, both could be dependent variables changing in response to changes in an unmeasured independent.

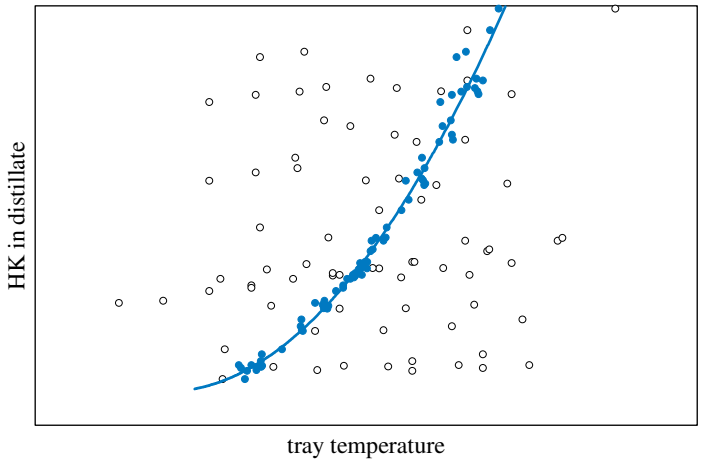

Figure 14.26 *Compensating an independent for variation in others*

The formulae above were developed assuming that we would wish to minimise, in the y direction, the sum of the squares of the distances from each point to the regressed line. If ε is the unaccounted error, then each value of y is given by

$$y_i = a_0 + a_1 x_i + \varepsilon_i \tag{14.118}$$

This assumes that the error arises in y, for example, because there are unmeasured variables affecting its value. However the error may be present in x, for example, because it is inaccurately measured. In this case we should develop a correlation of the form

$$y_i = a_0 + a_1 \left(x_i + \varepsilon_i \right) \tag{14.119}$$

In other words, we would wish to minimise

$$\sum_{i=1}^{n} \left(\hat{x}_i - x_i \right)^2 = \frac{1}{a_1^2} \sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2 \tag{14.120}$$

Alternatively x can be regressed against y to give the relationship

$$\hat{x} = b_0 + b_1 y \tag{14.121}$$

Inverting this equation gives

$$\hat{y} = a_0 + a_1 x \qquad \text{where} \qquad a_0 = -\frac{b_0}{b_1} \quad \text{and} \quad a_1 = \frac{1}{b_1} \tag{14.122}$$

We should adopt this approach if the error in x is large compared to the range of values of x. However, in practice there are likely to be measurement errors in both x and y. It is possible to give more equal weighting to both by using instead the perpendicular distance from each point to the regressed line. The function we would then minimise is given by

$$\sum_{i=1}^{n} \left(\hat{x}_i - x_i \right)^2 + \sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2 = \left(\frac{1}{a_1^2} + 1 \right) \sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2 \tag{14.123}$$

It can be shown that the value of a_1 derived using this penalty is the geometric mean of the values derived using the penalties described by Equations (14.90) and (14.120).

An alternative method of giving equal weight to potential errors in both x and y is, for each point, to multiply the distances from the line in both the horizontal and vertical directions. This is the area of the rectangle with one vertex at the predicted value and the opposite vertex at the true value. The penalty function is then the total area of all the rectangles and becomes

$$\sum_{i=1}^{n}\left|\hat{x}_i - x_i\right|.\left|\hat{y}_i - y_i\right| = \frac{1}{|a_1|}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \tag{14.124}$$

Using the data included in Table 14.3, the impact on the resulting correlation between %C4 and tray 17 temperature of using each of the penalty functions is given in Table 14.4. Because there is only one independent, each method gives the same result for Pearson R^2 – no matter which coefficients are chosen. The parameter (ϕ), recommended in Chapter 9, measures the accuracy of predicting y and so, not surprisingly, is worse for the correlation derived for accurately predicting x. Similarly the two methods giving equal weight to errors in both x and y, as might be expected, give intermediate values for ϕ. In this case, all the methods give similar coefficients and the difference in their accuracy is unlikely to be noticeable in view of other potential sources of error. The high value of Pearson R^2 reflects that there appears to be little error in the measurement of both x and y. Applying the same techniques to less accurate data would result in significantly different estimates for a_1. For example, Figure 14.27 shows the correlations developed using data with a Pearson R^2 of 0.722. The value of a_1 varies substantially from −0.564 for the conventional penalty function to −0.781 for that based on Equation (14.120). Remembering this is likely to be a process gain, the variation of around ±16% is substantial. As expected the functions described by Equations (14.123) and (14.124) generate intermediate values.

The formulae presented above can each be expanded to develop inferentials involving more than one independent variable. Additionally, weighting factors can be included that reflect the relative accuracy of each variable. For example, we apply the weighting w_0 to the dependent variable, w_1 to the independent variable x_1 and w_2 to x_2, etc. The penalty function then becomes

Table 14.4 *Comparison of penalty functions*

Function	a_0	a_1	R^2	Φ
$\sum_{i=1}^{n}\left(a_0 + a_1x_i - y_i\right)^2$	−28.09	0.5494	0.978	0.978
$\frac{1}{a_1^2}\sum_{i=1}^{n}\left(a_0 + a_1x_i - y_i\right)^2$	−28.65	0.5587	0.978	0.977
$\left(\frac{1}{a_1^2}+1\right)\sum_{i=1}^{n}\left(a_0 + a_1x_i - y_i\right)^2$	−28.46	0.5555	0.978	0.960
$\frac{1}{\lvert a_1\rvert}\sum_{i=1}^{n}\left(a_0 + a_1x_i - y_i\right)^2$	−28.83	0.5617	0.978	0.939

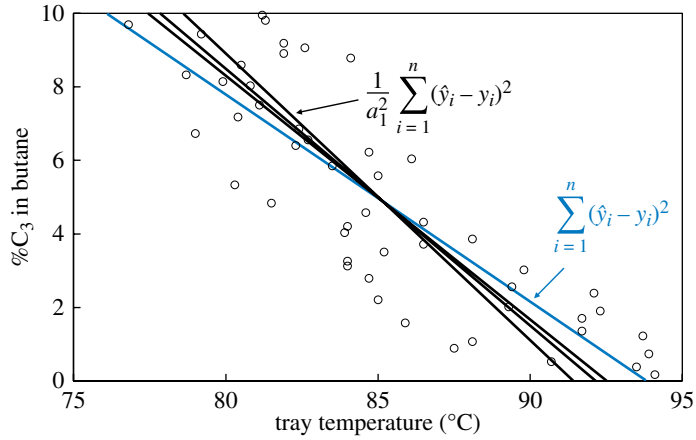

Figure 14.27 *Impact of choice of penalty function*

$$\sum_{i=1}^{n}\left[w_0\left(\hat{y}_i-y_i\right)^2+w_1\left(\hat{x}_{1i}-x_{ii}\right)^2+w_2\left(\hat{x}_{2i}-x_{2i}\right)^2+\ldots\right]$$
$$=\left(w_0+\frac{w_1}{a_1^2}+\frac{w_2}{a_2^2}+\ldots\right)\sum_{i=1}^{n}\left(\hat{y}_i-y_i\right)^2 \qquad (14.125)$$

The penalty functions described by Equations (14.90), (14.120) and (14.123) can all be developed from this equation by choosing appropriate weighting factors. It should be noted that selection of the best weighting factors cannot be achieved by choosing those that give the highest value of Pearson R^2 or ϕ. Both R^2 and ϕ are based on only the prediction error in y, and so will always be maximised by setting w_0 to 1 and all the other weighting factors to 0. Weighting coefficients can only be selected using judgement as to which measurements the engineer believes are more reliable. Avoiding having to make this judgement provides additional motivation to ensure all the data used are as reliable as possible. The more complex penalty functions would then offer no advantage over using the simplest conventional approach.

Equations could be identified for regressing nonlinear functions, for example

$$\hat{y}=a_0+a_1x+a_2x^2 \qquad (14.126)$$

However, a simpler approach is to treat x and x^2 as independent variables; x_1 is defined as x and x_2 as x^2. While x_1 and x_2 will now be correlated, the relationship between them is nonlinear and will not therefore cause problems in solving the equations. It is generally not advisable to exclude the linear term when adding the quadratic term. By differentiating Equation (14.126) we can determine at what value of x the minimum (or maximum) occurs for the predicted value.

$$\frac{d\hat{y}}{dx}=a_1+2a_2x=0 \qquad \text{hence} \qquad x=-\frac{a_1}{2a_2} \qquad (14.127)$$

By excluding the linear term we force the minimum (or maximum) to occur at $x = 0$. This is similar, when fitting a linear correlation, to omitting the constant term (a_0) and so force the predicted value to be 0 when $x = 0$.

An alternative is to apply some transformation to the predicted value, for example

$$\sqrt{\hat{y}}=a_0+a_1x \qquad (14.128)$$

is effectively the same as Equation (14.126) – although the values of a_0 and a_1 will be different. Other transformations such as the logarithm of the predicted value, or its square, might also be considered.

It may be beneficial to include *compound* inputs – those derived from a calculation involving several inputs. For example, the coefficients in an inferential may need to be adjusted if the type of feed being processed is changed. So, for Type 1 feed, the correlation might be

$$\hat{y} = a_0 + a_1 x_1 \tag{14.129}$$

For Type 2 it might be

$$\hat{y} = b_0 + b_1 x_1 \tag{14.130}$$

To ensure the same inferential can be used all the time we can incorporate the feed type by adding a second input (x_2) which is set to 1 when the feed is Type 1 and set to 0 when it is Type 2. So

$$\hat{y} = x_2 \left(a_0 + a_1 x_1\right) + \left(1 - x_2\right)\left(b_0 + b_1 x_1\right) \tag{14.131}$$

Rewriting

$$\hat{y} = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_1 x_2 \tag{14.132}$$

where

$$c_0 = b_0 \qquad c_1 = b_1 \qquad c_2 = a_0 - b_0 \qquad c_3 = a_1 - b_1 \tag{14.133}$$

Now imagine that the unit has never processed 100% of either feed type. We therefore do not have any data to regress to determine the coefficients in Equations (14.129) and (14.130). But provided we know x_2 (the proportion of Type 1 in the total feed) we can regress to determine directly the coefficients in Equation (14.132). However for this to be successful, if a_1 is significantly different from b_1, we must include the compound input $x_1 x_2$. The coefficient of x_1 is now effectively $(c_1 + c_3 x_2)$; or we can think of the coefficient of x_2 being effectively $(c_2 + c_3 x_1)$. Such functions, in which the coefficient of one input changes as another input varies, are known as *heterogenic*. (Those in which this is not the case are *homogenic.*) For example, as described in Figure 12.109, the coefficient used in a linear pressure compensated temperature varies with pressure and so including the compound variable *P.T* in a product composition inferential can improve accuracy.

The regression technique so far described is known as *ordinary least squares (OLS)*. OLS assumes that the data is *homoscedastic*, i.e. each data point is equally reliable. A refinement is *weighted least squares (WLS)* where weights are assigned to individual data points. For example a lower weighting might be applied to an input when its value falls within a range over which its accuracy is known to be suspect. An indication of this might be a correlation between the variance of the prediction error and one of the inputs. For example, using the data presented in Table 14.5, applying OLS gives the inferential for $\%C_4$ in distillate based on PCT (pressure compensated temperature).

$$\hat{y} = -12.09 + 0.3031x \tag{14.134}$$

For convenience the table has been sorted by increasing PCT. By separately calculating variance of the prediction error for each of the 5°C intervals, we can see that the reliability of the prediction varies greatly depending on the value of the PCT. We thus choose a weight (w) for each interval which, in this case, is the reciprocal of the variance of the error. We then modify Equation (14.90) to convert it to a weighted penalty function

$$\sum_{i=1}^{n} w_i \left(\hat{y}_i - y_i\right)^2 = \sum_{i=1}^{n} w_i \left(a_0 + a_1 x_i - y_i\right)^2 \tag{14.135}$$

Table 14.5 *Use of weighted least squares regression*

PCT (x_i)	Measured $\%C_4$ (y_i)	Predicted $\%C_4$ ($\hat{y}_i$)	$(\hat{y}_i - y_i)^2$	$\frac{\sum(\hat{y}_i - y_i)^2}{n}$	w	Weighted $\%C_4$ prediction
40.3	0.39	0.05	0.1140	0.4346	2.301	0.23
42.3	1.89	0.65	1.5252			0.82
43.9	1.42	1.14	0.0800			1.29
44.3	1.12	1.26	0.0189			1.41
45.9	1.27	1.74	0.2206	1.1766	0.850	1.87
46.3	3.32	1.86	2.1310			1.99
47.0	1.84	2.07	0.0534			2.20
47.9	1.73	2.34	0.3749			2.46
48.0	1.73	2.37	0.4127			2.49
48.3	2.74	2.46	0.0768			2.58
49.9	1.53	2.94	2.0019			3.05
50.0	0.94	2.98	4.1412			3.08
50.3	3.57	3.07	0.2546	1.4968	0.668	3.16
51.9	2.59	3.55	0.9167			3.63
52.0	3.02	3.58	0.3109			3.66
52.3	5.27	3.67	2.5664			3.75
53.9	3.28	4.15	0.7570			4.22
54.0	3.22	4.18	0.9220			4.25
54.3	6.45	4.27	4.7498			4.34
55.9	4.64	4.75	0.0127	3.5368	0.283	4.80
56.0	4.09	4.78	0.4800			4.83
56.3	6.92	4.87	4.1894			4.92
57.9	5.42	5.36	0.0042			5.39
58.0	4.78	5.39	0.3665			5.42
58.1	2.23	5.42	10.1476			5.45
58.3	8.37	5.48	8.3765			5.51
59.3	9.11	5.78	11.1083			5.80
59.9	6.33	5.96	0.1385			5.98
60.0	5.25	5.99	0.5446			6.01

(*Continued*)

Table 14.5 (*Continued*)

PCT (x_i)	Measured $\%C_4$ (y_i)	Predicted $\%C_4$ ($\hat{y}_i$)	$(\hat{y}_i - y_i)^2$	$\frac{\Sigma(\hat{y}_i - y_i)^2}{n}$	w	Weighted $\%C_4$ prediction
60.1	3.87	6.02	4.6145			6.04
61.9	9.83	6.56	10.6898			6.56
62.0	6.04	6.59	0.3032	5.0110	0.200	6.59
62.1	4.13	6.62	6.2037			6.62
64.0	7.43	7.19	0.0561			7.18
64.1	4.36	7.22	8.1987			7.21
66.0	9.98	7.80	4.7707			7.76
66.1	6.52	7.83	1.7055	1.9080	0.524	7.79
68.1	8.70	8.43	0.0737			8.38
69.1	9.77	8.73	1.0820			8.67

Minimising this results in a modified inferential

$$\hat{y} = -11.58 + 0.2930x \tag{14.136}$$

The weighted error variances for each 5°C interval will now be approximately the same. Whether the resulting inferential is any more reliable is debatable. In practice choosing the weights, without a strong engineering basis for their inclusion, will add nothing to the accuracy of the resulting correlation.

A better approach would be to explore what is causing the variation in accuracy. For example, in this case, it might be that the PCT calculation is not reliable over the whole range. If so, then correcting this would be a far better solution. It does however demonstrate that checking how the prediction error varies can highlight an opportunity to improve an inferential that may not be obvious otherwise. Figure 14.28, for example, shows the same correlation as Figure 14.26 but with the prediction error plotted against the actual dependent value. The prediction is based on a linear function of pressure and temperature. The curve crossing the zero error line in more than one place would suggest that the relationship is nonlinear. Indeed, Figure 14.26 revealed that a nonlinear PCT should be used – as described in Chapter 12. Other plots of prediction error can be helpful. For example plotting it against an independent not included in the correlation would, if a trend appears, suggest the independent should be included.

The *partial least squares (PLS)* method is applied in situations where the number of inputs exceeds the number of data sets. If the number of inputs is equal to the number of data sets then, in general, an exact fit to the data will be possible. Increasing the number of inputs beyond this will cause the technique covered in this section to fail. This is unlikely to arise for inferentials that the control engineer is likely to

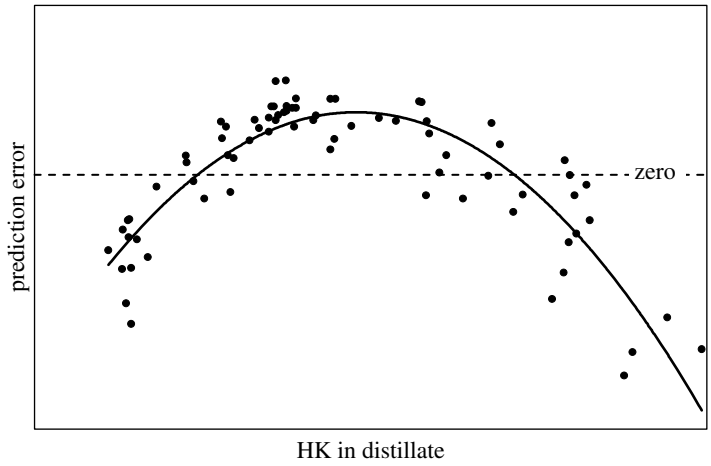

Figure 14.28 *Detecting whether a nonlinear correlation should be used*

develop. But it does arise, for example, with near infra-red (NIR) analysers. For calibration each sample these generate a frequency spectrum which will contain the amplitudes of around 1,000 frequencies. Only a few properties are predicted from this result. It would be impractical to collect the many thousands of samples that would be needed to apply OLS. Instead, PLS is used to first search for *latent vectors* that explain as much as possible of the relationship between the dependent properties and the independent amplitudes. These vectors are then used in developing the correlation. The mathematics are quite complex, well beyond the scope of this book, and in any case would be built into a software tool.

14.18 Outliers

Deciding which data points to exclude from regression analysis is very subjective. If the scatter is poor then the position of a single outlier will have a significant impact on the slope of the line of best fit. Basing any conclusion on the effect of a single point might be unwise and the point is better treated as an outlier. However, if there is evidence that the point is representative, excluding it would also result in a suspect conclusion.

The most frequently published method of defining outliers is to treat as *mild* any value which falls outside 1.5 times the interquartile range and as *extreme* any value outside three times the range. By definition there is a 50% probability of a value lying in the interquartile range. From Table 14.1 we can see that this corresponds to a range of $\pm 0.6745\sigma$. Multiplying this by 1.5 gives a value close to σ and by 3 a value close to 2σ. We could therefore adopt a policy of excluding all values more than 2σ from the mean and reviewing those that lie between σ and 2σ away from the mean. We are effectively choosing to exclude any value which we are 95% certain is an outlier.

Excluding outliers will change the mean and reduce the standard deviation of the values retained. Thus data points which were mild outliers may now fail the 2σ criterion. Excluding these, and any others that arise from the next recalculation of μ and σ, could severely undermine the confidence we have in the regression.

14.19 Model Identification

We can apply regression analysis to identify dynamic models. In general these models are based on discrete data collected at a fixed time interval (*ts*). They have the form

$$PV_i = a_0 + a_1 PV_{i-1} + a_2 PV_{i-2} \ \ldots + b_0 MV_i + b_1 MV_{i-1} \ \ldots + b_{i-\theta/ts} MV_{i-\theta/ts} \tag{14.137}$$

Models which use only previous values of the parameter they aim to predict are known as *autoregressive (AR) models*. Since the dynamic model includes values for *MV* in addition to previous values of *PV*, it is known as an *autoregressive with exogenous input (ARX) model*. These models are often used in MPC, as described in Chapter 8. We saw in Chapter 2 how the coefficients in this form of model could be used to derive the more conventional FOPDT constants (K_p, θ and τ) used in a parametric model. More complex equations can be used to identify higher order models. For example, in Laplace form, Equation (14.138) is that of a second order process (with lags τ_1 and τ_2, and lead τ_3).

$$\frac{PV}{MV} = \frac{K_p e^{-\theta s}\left(\tau_3 s + 1\right)}{\left(\tau_1 s + 1\right)\left(\tau_2 s + 1\right)} \tag{14.138}$$

Lead is required if there is PV overshoot ($\tau_3 > 0$) or inverse response ($\tau_3 < 0$). In Chapter 15 we will cover applying z-transforms to Laplace forms to generate *finite difference equations*. Doing so here gives

$$PV_n = a_1 PV_{n-1} + a_2 PV_{n-2} + b_1 MV_{n-\theta/ts} + b_2 MV_{n-\theta/ts-1} + bias \tag{14.139}$$

$$a_1 = e^{-ts/\tau_1} + e^{-ts/\tau_2} \qquad \text{and} \qquad a_2 = -e^{-ts/\tau_1} e^{-ts/\tau_2} \tag{14.140}$$

$$b_1 = K_p\left[1 + \frac{\tau_3 - \tau_1}{\tau_1 - \tau_2} e^{-ts/\tau_1} + \frac{\tau_2 - \tau_3}{\tau_1 - \tau_2} e^{-ts/\tau_2}\right] \tag{14.141}$$

$$b_2 = K_p\left[e^{-ts/\tau_1} e^{-ts/\tau_2} + \frac{\tau_2 - \tau_3}{\tau_1 - \tau_2} e^{-ts/\tau_1} + \frac{\tau_3 - \tau_1}{\tau_1 - \tau_2} e^{-ts/\tau_2}\right] \tag{14.142}$$

$$MV_{n-\theta/ts} = MV_{n-\text{int}(\theta/ts)} - \left(\frac{\theta}{ts} - \text{int}\left(\frac{\theta}{ts}\right)\right)\left(MV_{n-\text{int}(\theta/ts)} - MV_{n-\text{int}(\theta/ts)-1}\right) \tag{14.143}$$

$$MV_{n-\theta/ts-1} = MV_{n-\text{int}(\theta/ts)-1} - \left(\frac{\theta}{ts} - \text{int}\left(\frac{\theta}{ts}\right)\right)\left(MV_{n-\text{int}(\theta/ts)-1} - MV_{n-\text{int}(\theta/ts)-2}\right) \tag{14.144}$$

K_p, θ, τ_1, τ_2, τ_3 and *bias* can be fitted directly. Or an approach similar to that used for the first order process can be taken where a_1, a_2, c_1, c_2, c_3 and *bias* in Equation (14.145) are identified, for the best integer value of θ/ts, using linear regression.

$$PV_n = a_1 PV_{n-1} + a_2 PV_{n-2} + c_1 MV_{n-\theta/ts} + c_2 MV_{n-\theta/ts-1} + c_3 MV_{n-\theta/ts-2} + bias \tag{14.145}$$

The deadtime is then determined from the best integer value of θ/ts by

$$\theta = \left[\text{int}\left(\frac{\theta}{ts}\right) + \frac{c_2}{2c_1 + c_2}\right] ts \tag{14.146}$$

The coefficients b_1 and b_2 used in Equation (14.139) can be determined from

$$b_1 = \frac{2c_1 + c_2}{2} \tag{14.147}$$

$$b_2 = \frac{c_2\left(2c_1 + c_2\right)}{4c_1} \tag{14.148}$$

K_p can then be derived from

$$K_p = \frac{c_1 + c_2 + c_3}{1 - a_1 - a_2} = \frac{b_1 + b_2}{1 - a_1 - a_2} \tag{14.149}$$

The values for τ_1 and τ_2 are interchangeable. Arbitrarily we select $\tau_1 > \tau_2$, and so they can be derived from

$$\tau_1 = \frac{ts}{\ln\left(\dfrac{2}{a_1 + \sqrt{a_1^2 + 4a_2}}\right)} \tag{14.150}$$

$$\tau_2 = \frac{ts}{\ln\left(\dfrac{2}{a_1 - \sqrt{a_1^2 + 4a_2}}\right)} \tag{14.151}$$

To be solvable for τ_2, a_2 must be negative. In addition, for τ_1 and τ_2 to be real

$$a_1^2 + 4a_2 \geq 0 \tag{14.152}$$

In the process industry this will generally be the case. Processes where this condition is not met are described as underdamped. As described in Section 3.9 these are very rare and so, because of the more complex mathematics involved, their description is deferred to Chapter 15.

The value for τ_3 is obtained by substituting the results for K_p, τ_1 and τ_2 into either Equation (14.141) or Equation (14.142). For example, from Equation (14.141)

$$\tau_3 = \frac{\left(\dfrac{b_1}{K_p} - 1\right)\left(\tau_1 - \tau_2\right) + \tau_1 e^{-ts/\tau_1} - \tau_2 e^{-ts/\tau_2}}{e^{-ts/\tau_1} - e^{-ts/\tau_2}} \tag{14.153}$$

Note that Equations (14.141) and (14.142) cannot be applied if

$$a_1^2 + 4a_2 = 0 \tag{14.154}$$

We will see later that this is described as a critically damped process. It means that $\tau_1 = \tau_2$ and b_1 and b_2 would therefore be indeterminate. If $\tau_1 = \tau_2 = \tau$, then the coefficients become

$$a_1 = 2e^{-ts/\tau} \qquad \text{and} \qquad a_2 = -e^{-2ts/\tau} \tag{14.155}$$

$$b_1 = K_p\left[1 - e^{-ts/\tau} + \frac{(\tau_3 - \tau)ts}{\tau^2}e^{-ts/\tau}\right] \tag{14.156}$$

$$b_2 = K_p\left[e^{-2ts/\tau} - e^{-ts/\tau} - \frac{(\tau_3 - \tau)ts}{\tau^2}e^{-ts/\tau}\right] \tag{14.157}$$

Equation (14.149) can still be used to obtain K_p but the lags (τ) are obtained from Equation (14.155)

$$\tau = \frac{-ts}{\ln\left(\frac{a_1}{2}\right)} = \frac{-2ts}{\ln(-a_2)} \tag{14.158}$$

The value for τ_3 is obtained by substituting the results for K_p and τ into either Equation (14.156) or (14.157). For example, from Equation (14.156)

$$\tau_3 = \tau + \frac{\left(\frac{b_1}{K_p} - 1 + e^{-ts/\tau}\right)\tau^2}{e^{-ts/\tau}} \tag{14.159}$$

In Chapter 2 we showed an alternative derivation for a first order process, with a lag of τ_1, based on the differential equation

$$\tau_1\frac{dPV}{dt} + PV = K_pMV \tag{14.160}$$

If the PV is passed through a second lag (τ_2)

$$\tau_2\frac{d}{dt}\left(\tau_1\frac{dPV}{dt} + PV\right) + \tau_1\frac{dPV}{dt} + PV = K_pMV \tag{14.161}$$

This, on rearranging, gives us the differential equation for a simple second order process (excluding PV overshoot or inverse response).

$$\tau_1\tau_2\frac{d^2PV}{dt^2} + (\tau_1 + \tau_2)\frac{dPV}{dt} + PV = K_pMV \tag{14.162}$$

Expressing this as its discrete approximation gives

$$\frac{\tau_1\tau_2}{ts^2}(PV_n - 2PV_{n-1} + PV_{n-2}) + \frac{\tau_1 + \tau_2}{ts}(PV_{n-1} - PV_{n-2}) + PV_{n-2} = K_pMV_{n-\theta/ts} \tag{14.163}$$

This can be written in the same form as Equation (14.137) where

$$a_1 = 2 - \frac{\tau_1 + \tau_2}{\tau_1\tau_2}ts \tag{14.164}$$

$$a_2 = -1 + \frac{\tau_1 + \tau_2}{\tau_1\tau_2}ts - \frac{ts^2}{\tau_1\tau_2} \tag{14.165}$$

$$b_1 = \frac{K_p ts^2}{\tau_1 \tau_2} \tag{14.166}$$

Solving these gives

$$K_p = \frac{b_1}{1 - a_1 - a_2} \tag{14.167}$$

$$\tau_1 = \frac{2 - a_1 + \sqrt{a_1^2 + 4a_2}}{2\left(1 - a_1 - a_2\right)} ts \tag{14.168}$$

$$\tau_2 = \frac{2 - a_1 - \sqrt{a_1^2 + 4a_2}}{2\left(1 - a_1 - a_2\right)} ts \tag{14.169}$$

While strictly not as accurate, these last two equations will give much the same result as Equations (14.150) and (14.151).

We will show, in Chapter 15, that a second order process can also be described by defining a single lag, known as the *natural period of oscillation (τ)* and the *damping ratio (ζ)*

$$\tau^2 \frac{d^2 PV}{dt^2} + 2\zeta\tau \frac{dPV}{dt} + PV = K_p MV \tag{14.170}$$

where τ and ζ can then be derived from

$$\tau = \frac{ts}{\sqrt{1 - a_1 - a_2}} \tag{14.171}$$

$$\zeta = \frac{2 - a_1}{2\sqrt{1 - a_1 - a_2}} \tag{14.172}$$

Conversion (using either method) to parametric form however, is not necessarily a prerequisite for using an ARX model for control design. For example, controller tuning software can be designed to use either form. Indeed, the use of ARX models may be preferable. For example there is no need to distinguish between self-regulating and integrating processes. In analog form an integrating process is normally described by

$$\frac{dPV}{dt} = K_p . MV + bias \tag{14.173}$$

In digital form this becomes

$$\frac{PV_n - PV_{n-1}}{ts} = K_p . MV_{n-\theta/ts} + bias \tag{14.174}$$

Or

$$PV_n = aPV_{n-1} + bMV_{n-\theta/ts} + bias \qquad \left(a = 1 \quad \text{and} \quad b = K_p ts\right) \tag{14.175}$$

Another consideration is that, given the freedom, regression analysis will not often result in the selection of three consecutive historical values of *MV*, as suggested by Equation (14.145). Consider, as an example, the use of the difference between two distillation column tray temperatures as a measure of separation. Such a measurement is likely to be the PV of a controller intended to manipulate reboiler duty. Performing a step-test on reboiler duty in order to determine the process dynamics is likely to result in such a model. This is because the temperature difference will change first due to the change in tray temperature nearer the reboiler. The difference will change again some time later when the other tray temperature responds. Conversion of the resulting model to parametric form is unlikely to be practical.

Fitting higher order models is possible by increasing the number of historical values of PV and MV used in developing the correlation. While this will always increase the accuracy of the model it can result in over-fitting, so that any noise present will also be modelled. There are several techniques, described earlier in this chapter, which can be applied to determine the best choice of order.

ARX models can be further extended to become *autoregressive moving average with exogenous input (ARMAX)* models. The aim of the moving average is to adjust the model prediction to take account of other changes to the process – often misleadingly described as *noise*. The model includes another input which can be any measurement. For example, as Equation (14.176) shows, this might be a DV so that the model will then also predict the impact of a load change.

$$PV = \sum_{i=1}^{n} a_i PV_{n-i} + \sum_{j=1}^{p} b_j MV_{p-\theta/ts-j+1} + \sum_{k=1}^{q} c_k DV_{q-\theta/ts-k+1} + bias \tag{14.176}$$

In addition to linearly regressed models it is possible to apply nonlinear techniques, such as an artificial neural network described later in this chapter. A nonlinear ARX model, for instance, would then be described as a *NARX* (or *NLARX*) model.

14.20 Autocorrelation and Autocovariance

Autocorrelation is a correlation of a set of values with itself, sometimes also called *serial correlation*. External inputs are not considered. It is this type of model we explore here.

We apply Equation (14.25) to check for a correlation between a value measured now and the same measurement taken earlier, where k is the number of collection intervals between the two.

$$R_k = \frac{\sum_{i=k+1}^{n} \left(x_i - \bar{x}\right)\left(x_{i-k} - \bar{x}\right)}{\sqrt{\sum_{i=1}^{n-k} \left(x_i - \bar{x}\right)^2 \sum_{i=k+1}^{n} \left(x_{i-k} - \bar{x}\right)^2}} \tag{14.177}$$

One use is to identify if there is a repetitive pattern. For example Figure 14.29 shows a noisy measurement taken every second for 10 minutes. At first glance it might appear that the underlying measurement is constant. However, by determining the Pearson correlation coefficient (R) at different values of k, we can see from the *correlogram* (shown as Figure 14.30) that the correlation varies cyclically. The strongest positive correlation occurs at regular intervals of about 90 seconds, as does the strongest negative correlation. This tells us that the apparently constant measurement has an underlying oscillation with a period of about 3 minutes. This might highlight a previously undetected problem with controller tuning or an issue with the control valve. Indeed, this technique forms the basis of some controller oscillation detection

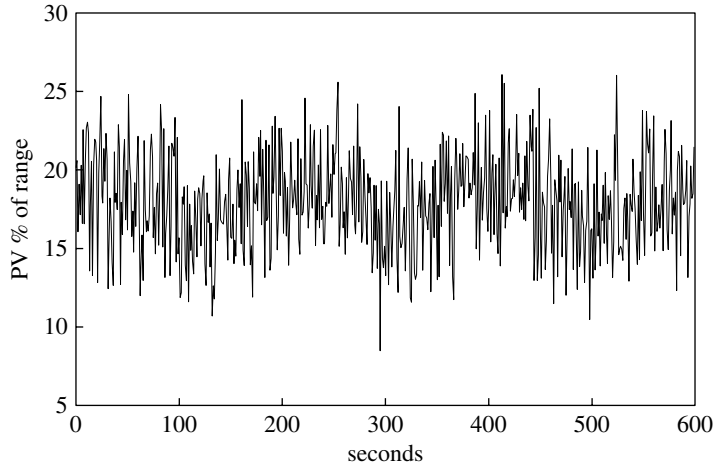

Figure 14.29 *Noisy measurement*

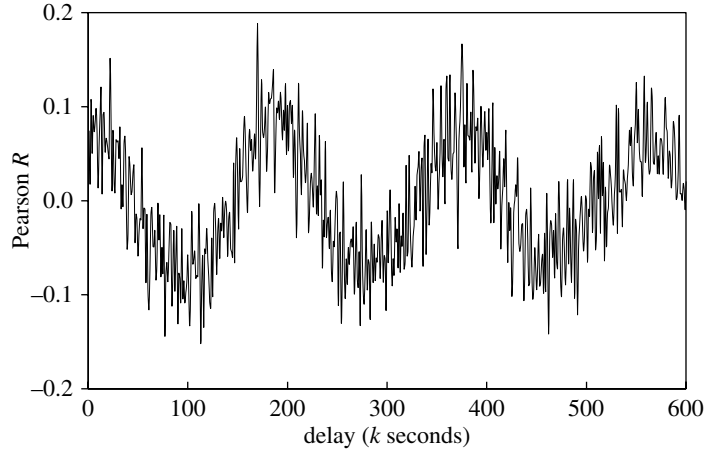

Figure 14.30 *Oscillation detected by autocorrelation*

techniques. The correlogram has the same frequency of oscillation as the raw measurement but with a much improved signal-to-noise ratio – making automatic detection of oscillation more reliable. The same underlying oscillation would also be detected by applying the *discrete Fourier transform* technique described later in Chapter 15.

While not strictly autocorrelation (because it involves another measure of the same parameter), it can be used to explore whether a correlation between two measurements requires timing to be taken into account. For example the engineer may wish to validate an inferential property by comparison with an on-stream analyser. The inferential property is likely to respond more quickly to process changes than the analyser. This is not an issue if comparison data are only collected when the process is at steady state. If this is not the case then the two measurements will not agree, potentially leading to the conclusion wrongly that one

is suspect. Again, plotting R_k against k will identify what delay to apply to the inferential measurement before comparing it to the analyser. Data collected under changing conditions can then be combined with data collected at steady state. This is illustrated by Figure 14.31. In response to a series of step-tests the inferential (coloured line) shows a deadtime of 0.5 minutes and a lag of 2 minutes. The analyser (black line) has a deadtime of 10 minutes and a lag of 0.5 minute. We could identify these models and apply full dynamic compensation to the inferential, as described in Chapter 9. But for simpler comparison, Figure 14.32 shows the optimum delay to apply would be 8.3 minutes. The dashed line in Figure 14.31 shows this applied to the inferential. Without compensation for the delay, the correlation coefficient (R) between analyser and inferential would be around 0.78 but, with it, it approaches a value of 1.

The same approach can be applied to the development of inferentials, checking whether a delay should be applied to any of the independent variables being considered as inputs to the inferential calculation. Similarly it can be applied to dynamic model identification, as described in Chapter 2, to determine a good starting estimate for process deadtime that is then improved by iteration. It can also be used to validate the dynamic models in MPC.

Also used is *autocovariance*, derived from Equation (14.22).

$$g_k = \frac{\sum_{i=k+1}^{n} \left(x_i - \bar{x}\right)\left(x_{i-k} - \bar{x}\right)}{n-k-1} \tag{14.178}$$

While similarly effective, like covariance, its magnitude depends not only on whether there is a correlation but also on the magnitude of the values x. While a nonzero value would indicate a correlation, it is not immediately obvious how strong it is.

An important use is in modelling process behaviour. Consider the inventory of liquid in a product storage tank. The inventory is clearly a function of the product flow into the tank but is also a function of liquid being pumped from the tank to another process or intermittently in batches for delivery to a customer. While not necessarily of obvious value to the control engineer, modelling stock levels can be an important part of feasibility studies. For example gasoline is a blend of naphtha type components and

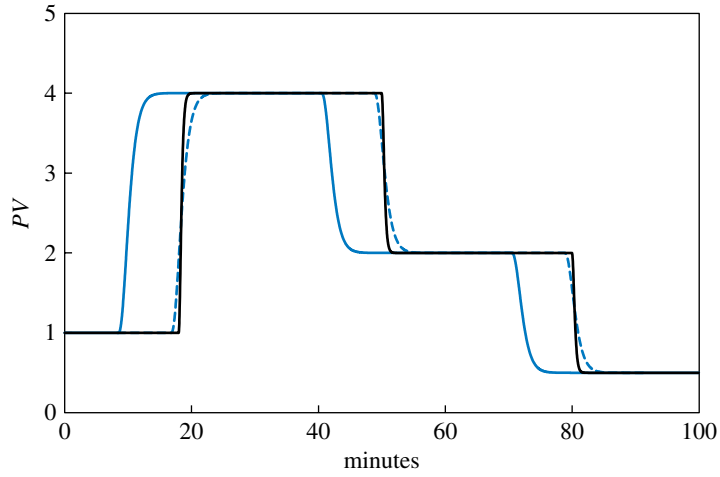

Figure 14.31 *Comparison of on-stream analyser with inferential*

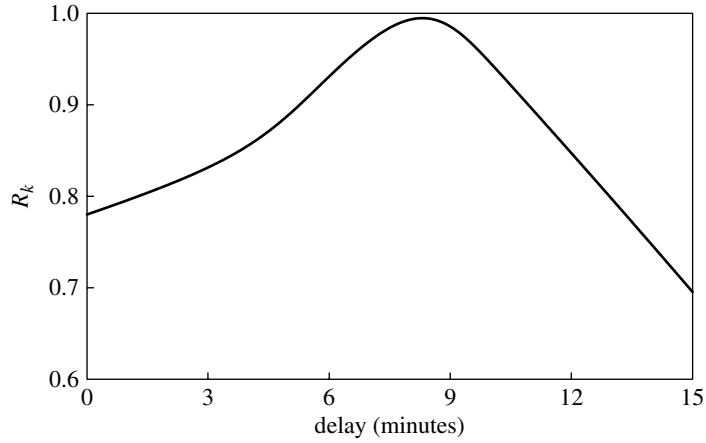

Figure 14.32 *Effect of time delay on correlation*

small amounts of butane. It is usually economic to maximise the butane content up to the RVP specification of the product. In industry there are many examples where the economic advantage of doing so has been overestimated because of periods when insufficient butane was available in storage. Being able to predict such occasions would be an important part of estimating the benefits before embarking on a blend optimisation project.

We are likely to have available inventory data collected regularly, usually at the end of each day. Analysing these data is likely to show that the inventory follows a normal distribution with mean and standard deviation that we can readily determine. However, we cannot use this distribution to generate inventories for our simulation. If a tank is almost full one day, it is very unlikely to be nearly empty the next. Today's inventory is a function of yesterday's inventory. As an example Table 14.6 and Figure 14.33 show the inventory recorded daily over 220 days.

Our data is a *time series* and needs to be modelled as such. If x is the inventory, a possible approach would be to develop a model such as

$$x_i = x_{i-1} + z_i \tag{14.179}$$

In this case z is a normally distributed random variable with a mean and standard deviation that we have identified by analysing the difference between x_{i-1} and x_i. Thus z is the change in inventory each day; this however, if selected from a normal distribution, could result in the predicted inventory being negative or greater than the tank capacity (in this case, 2258). Further, the changes in inventory may also be a time series. This is clear from Figure 14.34 which shows the cumulative sum of deviations from the mean inventory. If changes in inventory were truly random then we would expect the trend to be very noisy around zero, not to show long trends upward or downward. We might also expect cyclical patterns, for example, product withdrawal may be different at weekends and we would then expect x_i to be correlated with x_{i-7}.

Figure 14.35 shows how the autocorrelation coefficient (R^2) varies with the age of the previous inventory measurement. As expected there is a strong correlation with yesterday's level; the correlation then declines rapidly for older values. But, as we anticipated, there are peaks at one and two weeks. The peak at six weeks might be due to some other cyclical behaviour. It coincides with the changes in direction of

Table 14.6 *Tank inventories by day*

1–25	26–50	51–75	76–100	101–125	126–150	151–175	176–200	201–220
422	1118	620	1250	1485	435	1589	1785	560
520	1380	260	1510	1485	625	1470	1920	820
620	1740	420	1510	1485	940	1450	1510	1085
855	2075	620	1395	1485	1000	1680	1155	1365
970	1780	800	1295	1485	1025	2000	1480	1640
1210	1785	1050	1135	1485	960	1705	1780	1900
1475	1940	710	910	1175	1340	1255	2080	1470
1490	1930	1060	850	840	1780	1450	1680	1100
1560	1660	560	845	620	2000	1770	1570	1045
1718	1740	950	740	485	1910	2095	1790	1080
1837	1895	1220	745	445	1575	2090	2025	1410
1820	1785	1105	980	470	1205	1920	1795	1750
1860	1450	1105	1180	560	945	1600	1665	1640
1845	990	750	960	620	885	1630	1540	1320
1560	1260	435	925	680	1180	1950	1100	1365
1465	1490	75	960	700	1400	1860	940	1340
1520	1720	360	1060	455	1760	1685	1225	1360
1595	1880	560	1245	780	1770	1690	1560	1420
1500	1600	720	1480	1185	1140	2000	1820	1645
1400	1685	960	1480	1140	470	1485	1995	1850
1450	1495	950	1480	1070	505	1680	1605	
1740	1625	902	1480	700	845	2000	1310	
2040	1450	900	1480	300	1180	1590	960	
1900	1450	970	1480	280	1525	1440	690	
1500	1065	1095	1490	340	1865	1470	580	

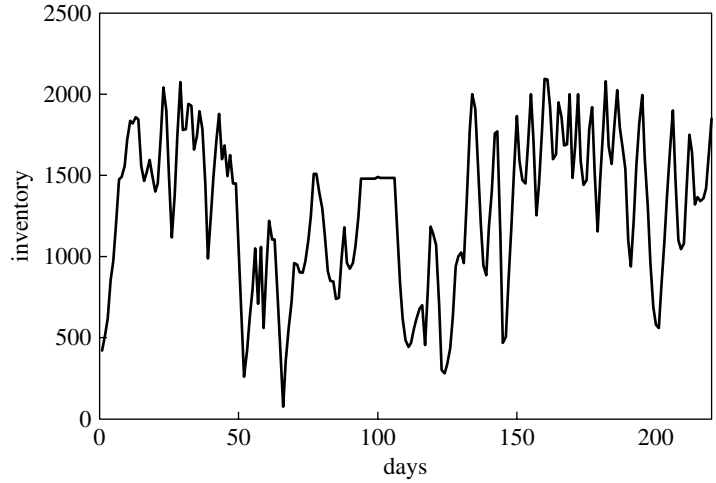

Figure 14.33 *Daily records of inventory*

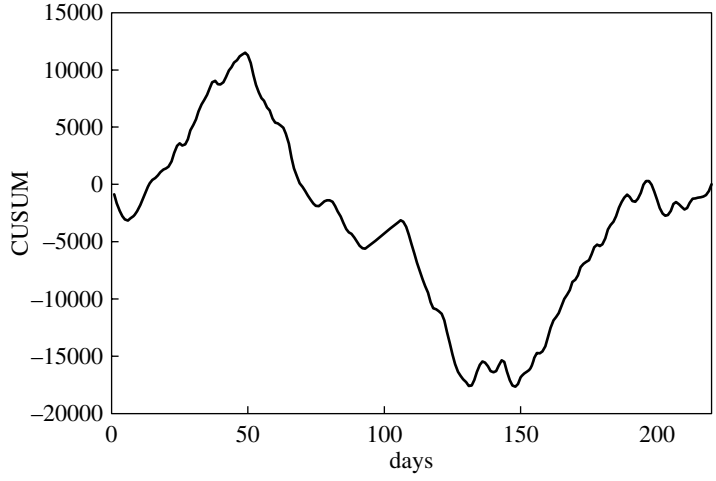

Figure 14.34 *Cumulative sum of deviations from mean*

the CUSUM trend and might result from a regular change in operating mode. Otherwise it would indicate quite an unlikely correlation. It should be included with care.

The general approach is to develop a model using the last p deviations from the mean to predict the next deviation.

$$x_i = \bar{x} - \sum_{j=1}^{p} b_j \left(x_{i-j} - \bar{x} \right) + z_i \tag{14.180}$$

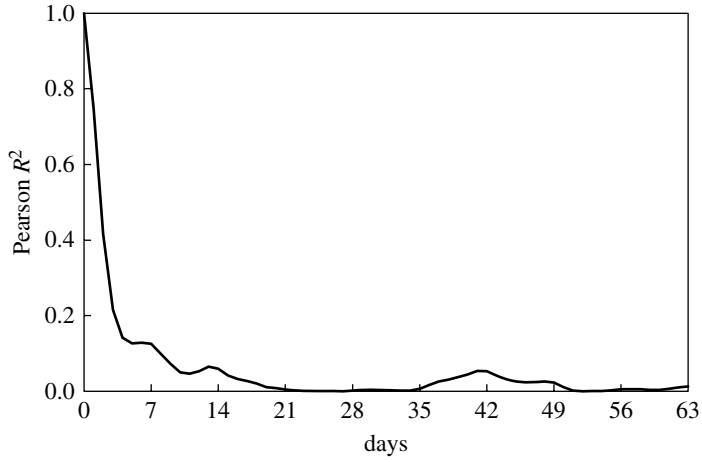

Figure 14.35 *Autocorrelation of inventories*

The coefficients (b) are chosen to give the best fit. The difference between the actual deviation and that predicted will be a random number (z_i) with a mean of zero and a known standard deviation. This variable is sometimes described as *white noise*.

To determine the coefficients (b) we choose an initial value for the number of days (p) used in the model. Using Equation (14.178) we then calculate p autocovariances (g_0 to g_p). Note that g_0 is the variance of x (σ^2). Using the method developed by Anderson [43], we can find the coefficients by solving the following simultaneous equations.

$$\begin{pmatrix} g_0 & g_1 & g_2 & \cdot & \cdot & g_{p-1} \\ g_1 & g_0 & g_1 & \cdot & \cdot & g_{p-2} \\ g_2 & g_1 & g_0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & g_2 \\ g_{p-2} & \cdot & \cdot & \cdot & g_0 & g_1 \\ g_{p-1} & g_{p-2} & \cdot & g_2 & g_1 & g_0 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \cdot \\ b_{p-1} \\ b_p \end{pmatrix} = \begin{pmatrix} -g_1 \\ -g_2 \\ -g_3 \\ \cdot \\ -g_{p-1} \\ -g_p \end{pmatrix} \tag{14.181}$$

The solution is given by multiplying this equation by the inverse of the autocovariance matrix. This is a relatively simple technique in most spreadsheet packages.

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \cdot \\ b_{p-1} \\ b_p \end{pmatrix} = \begin{pmatrix} g_0 & g_1 & g_2 & \cdot & \cdot & g_{p-1} \\ g_1 & g_0 & g_1 & \cdot & \cdot & g_{p-2} \\ g_2 & g_1 & g_0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & g_2 \\ g_{p-2} & \cdot & \cdot & \cdot & g_0 & g_1 \\ g_{p-1} & g_{p-2} & \cdot & g_2 & g_1 & g_0 \end{pmatrix}^{-1} \begin{pmatrix} -g_1 \\ -g_2 \\ -g_3 \\ \cdot \\ -g_{p-1} \\ -g_p \end{pmatrix} \tag{14.182}$$

The mean of the random variable (z) will be 0; its standard deviation given by

$$\sigma_z^2 = \frac{1}{n-p}\sum_{i=p+1}^{n}\left(\left(x_i - \bar{x}\right) + \sum_{j=1}^{p} b_j\left(x_{i-1} - \bar{x}\right)\right)^2 \tag{14.183}$$

To avoid over-fitting the model we check whether each coefficient is significantly different from 0. So we make the null hypothesis that $b_i = 0$. Using a_{ii}, the appropriate diagonal element from the inverted autocovariance matrix, we determine ε_i.

$$\varepsilon_i = \frac{b_i}{\sigma_z}\sqrt{\frac{n-p}{a_{ii}}} \tag{14.184}$$

The value ε_i will be normally distributed with a mean of 0 and a standard deviation of 1. If the absolute value of ε_i is greater than 1.64 (see Figure 13.11), then we are 95% sure that b_i should not be zero. If this test shows b_p that should be nonzero then it is likely that p should be increased before proceeding further. Once we are satisfied that we have included sufficient historical values, other coefficients failing this test should be set to 0. For the model to be stable

$$-1 < \sum_{i=1}^{p} b_i < 0 \tag{14.185}$$

The calculated values of b_i will always obey this; however, setting to 0 any coefficients found to be insignificant can cause the total of the remainder to lie outside this range. In any case, because they are not independent, the remaining coefficients should be recalculated. This is done by removing the corresponding rows and columns from the autocovariance matrix and repeating the calculations above. There is no guarantee that this will result in the remaining coefficients being significantly different from 0; so the process may need repeating.

The first estimate of the coefficients gave the results in Table 14.7. The analysis was initially performed by including the last 30 values in the model but the coefficients applied to values more than 14 days old were shown to be insignificant ($\varepsilon < 1.64$) and the remainder recalculated. It now shows that the coefficients b_3 to b_6 and b_9 to b_{12} should all be set to 0. Table 14.8 shows the impact of doing so and reveals that b_8 should also be zero. Table 14.9 shows the impact of removing this coefficient; all the remaining coefficients are significantly different from 0.

The mean of the inventories is 1292 and the standard deviation 467. From Equation (14.183) we find that the standard deviation of the prediction error (σ_z) is 206. Our model therefore accounts for about 56% of the variation in inventory.

To generate the remaining behaviour we randomly select, for each predicted value, v_1 to v_{12} from the range 0 to 1 and use these to generate z_i.

$$z_i = 206\left[\sum_{j=1}^{12} v_j - 6\right]_i \tag{14.186}$$

$$x_i = 1292 + 1.177x_{i-1} - 0.402x_{i-2} + 0.066x_{i-7} + 0.147x_{i-13} - 0.115x_{i-14} + z_i \tag{14.187}$$

Table 14.7 *Initial estimate of coefficients*

	b	ε
1	−1.231	18.10
2	0.508	4.73
3	−0.030	0.27
4	−0.049	0.44
5	−0.071	0.64
6	0.096	0.87
7	−0.197	1.79
8	0.198	1.79
9	−0.174	1.57
10	0.172	1.54
11	−0.146	1.30
12	0.132	1.17
13	−0.218	2.03
14	0.127	1.87
total	−0.833	

Table 14.8 *Second estimate of coefficients*

	b	ε
1	−1.178	18.88
2	0.397	6.40
7	−0.147	2.35
8	0.099	1.59
13	−0.149	2.40
14	0.111	1.78
total	−0.866	

Table 14.9 *Final estimate of coefficients*

	b	ε
1	−1.177	18.82
2	0.402	6.46
7	−0.066	1.84
13	−0.147	2.37
14	0.115	1.84
total	−0.873	

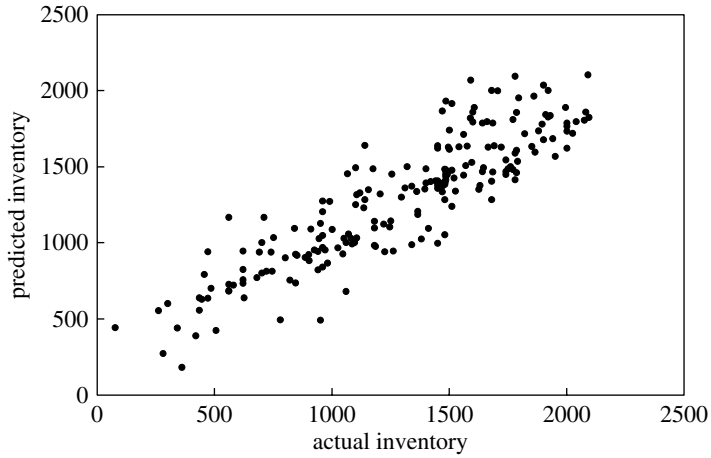

Figure 14.36 *Accuracy of predicted inventory*

Unlike process dynamic models, this model will not closely reproduce previous behaviour. In this example the daily inventory is only partially predictable; a significant component is noise caused by the semi-random nature of the process. Figure 14.36 shows its performance. While this might appear a poor model it does generate a time series with the same statistics as the actual data, and is therefore quite suitable for simulation studies.

14.21 Artificial Neural Networks

Artificial neural networks are developed using regression. Figure 14.37 shows the network as usually drawn. It comprises the *input layer* of *neurons*, each of which scale each of the n inputs using *weighting coefficients (w)* and a *bias (b)*.

$$x'_j = b_j + \sum_{i=1}^{n} w_{ij} x_i \tag{14.188}$$

The next *hidden layer* comprises the j neurons, in which each scaled input passes through a *transfer function* (also known as the *activation function*). This is chosen to be highly nonlinear with the output bounded within a narrow range. Most commonly used is the *sigmoid (S-shaped) function*, an example of which (shown as the coloured curve in Figure 14.38) is

$$y_j = \frac{1}{1+\exp\left(-x'_j\right)} \tag{14.189}$$

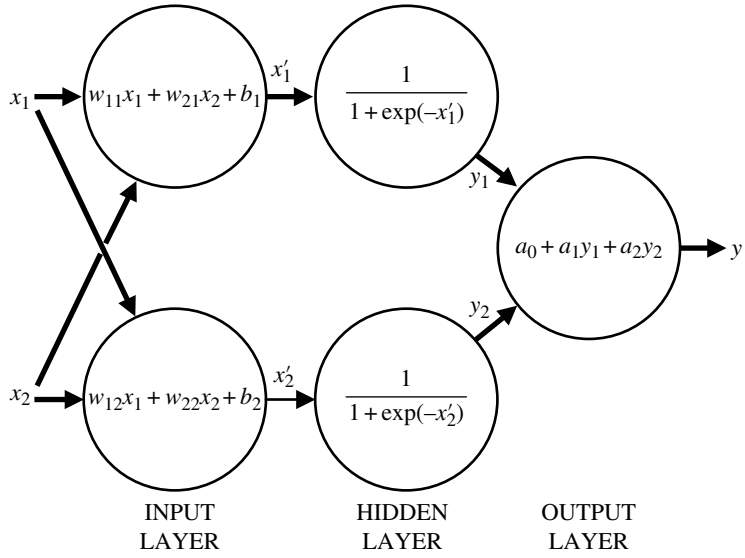

Figure 14.37 *Structure of artificial neural network*

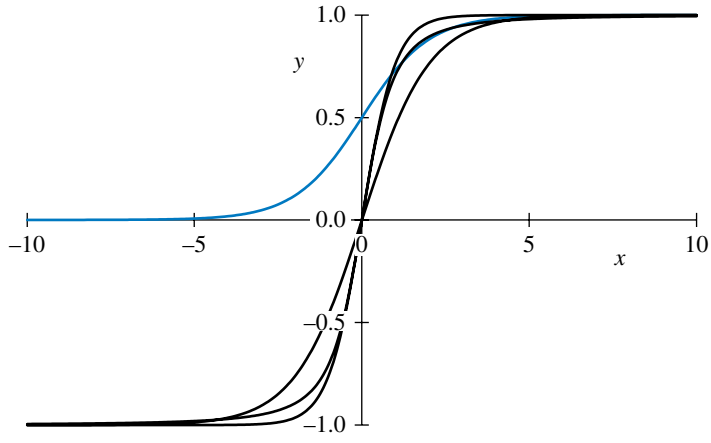

Figure 14.38 *Sigmoid curves*

where

$$-\infty < x'_j < \infty \qquad \text{and} \qquad 0 < y_j < 1 \tag{14.190}$$

There are other examples (shown as the black curves in Figure 14.38) where

$$-\infty < x'_j < \infty \qquad \text{and} \qquad -1 < y_j < 1 \tag{14.191}$$

such as

$$y_j = \frac{1-\exp\left(-x'_j\right)}{1+\exp\left(-x'_j\right)} \tag{14.192}$$

$$y_j = \tanh\left(x'_j\right) = \frac{\exp\left(x'_j\right)-\exp\left(-x'_j\right)}{\exp\left(x'_j\right)+\exp\left(-x'_j\right)} \tag{14.193}$$

$$y_j = \frac{x'_j}{\sqrt{1+x'^2_j}} \tag{14.194}$$

Others include the *radial basis function* (*RBF*); one example of which uses a Gaussian activation function based on the mean (μ) and standard deviation (σ) of x'_j.

$$y_j = \exp\left[-\frac{\left(x'_j-\mu_j\right)^2}{\sigma_j^2}\right] \tag{14.195}$$

There is also a *triangular basis function* (*TBF*).

$$y_j = 1-\left|x'_j\right| \quad \text{if } -1 < x'_j < 1; \quad y_j = 0 \quad \text{otherwise} \tag{14.196}$$

In the *output layer* of neurons the outputs from each transfer function are combined on a linear basis as a weighted sum

$$y = a_0 + \sum_{j=1}^{n} a_j y_j \tag{14.197}$$

Since neurons in one layer only pass information to all the neurons in the next layer, the network is of the *feedforward* type. In a *recurrent* network neurons also pass information to the previous layer. The coefficients (b, w and a) are chosen, as with linear regression, to minimise the residual sum of squares (*RSS*). With neural networks this process is iterative and is known as *training*.

Table 14.10 shows process data collected from the same distillation column as the example used in Section 14.12 except that it covers a wider range of operating conditions. Revising, by linear regression, the coefficients in Equation (14.56) will make it fit the process data reasonably well with *RSS* of 8.0.

$$y = -10.68 - 0.3619x_1 + 0.6668x_2 \tag{14.198}$$

Table 14.10 *Process data collected for development of inferential*

Tray 15 (x_1)	Tray 17 (x_2)	$\%C_4$ (y)
63.7	52.2	1.58
57.3	48.9	1.36
62.2	54.1	2.57
66.8	61.1	5.38
67.2	61.9	5.84
65.8	54.4	1.99
59.6	51.1	1.82
64.6	57.5	3.68
68.9	58.4	2.89
68.5	64.3	7.63
45.2	40.9	0.20
67.4	56.4	2.40
60.6	52.2	2.08
48.6	42.6	0.40
51.3	44.2	0.61
66.4	60.5	5.11
52.9	45.3	0.76
66.4	60.4	4.99
60.1	49.0	1.10
54.8	46.7	0.99
69.0	65.3	8.50
66.2	57.6	3.21
40.8	39.2	0.04
66.4	59.4	4.24
69.2	65.8	9.05
65.1	58.2	3.96
69.6	66.5	9.87
67.4	62.5	6.36
62.8	55.0	2.84
61.5	53.2	2.33

It is possible to develop a neural network with a single neuron in the hidden layer. The formula for this can be derived by combining Equations (14.188), (14.189) and (14.197).

$$y = a_0 + \frac{a_1}{1 - \exp\left[-b_1 - w_{11}x_1 - w_{21}x_2\right]} \tag{14.199}$$

Training results in

$$b_1 = -8.8896578 \qquad w_{11} = -0.0409135 \qquad w_{21} = 0.1101842 \tag{14.200}$$

$$a_0 = -1.3965854 \qquad a_1 = 919.7932 \tag{14.201}$$

This gives *RSS* a value of 0.8, considerably better than that from the linear regression. Figure 14.39 confirms that the predicted values now match the measured values very closely. While it would be possible to include nonlinear functions in conventional regression it is unlikely that the use of simple transformations, such a logarithm and powers, will match the *RSS* achieved by the neural network.

As discussed in Chapter 9, and earlier in this chapter, it is important that the inferential makes good engineering sense. With simple functions the meaning of unusual terms (for example x_1x_2) can be checked and the signs of their coefficients can be validated. With a neural network this is impractical. Indeed, if using a proprietary package, the values of the coefficients are not always made available to the engineer. However, we can explore how the inferential behaves by varying each of its inputs. Figure 14.40 shows the result of this exercise; the dashed lines compare the performance to the linear inferential. While analysis does not show that the inferential is accurate it demonstrates that it behaves as expected. The $\%C_4$ increases both as the tray 17 temperature increases and as the temperature difference between the trays falls. While this gives a strong indication that the inferential is reliable we would need to be cautious in applying it if any input falls outside the range over which training was performed.

It would be possible, in this example, to increase the number of neurons. While this would reduce the RSS, it would do so by modelling more of the noise in the measurements. This *over-fitting* would likely result in a reduction in accuracy when the inferential is put into use. One check is to look at the relative

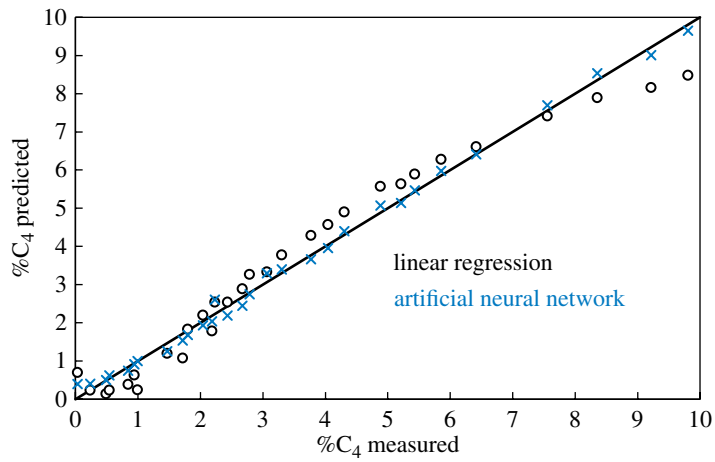

Figure 14.39 *Improved performance of inferential using artificial neural network*

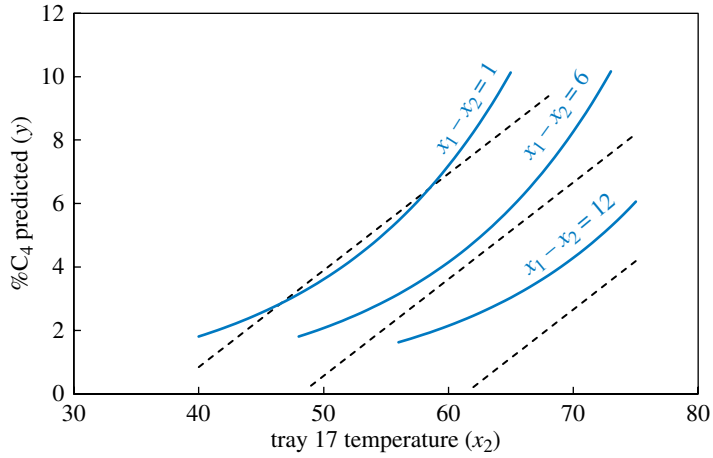

Figure 14.40 *Validating inferential based on artificial neural network*

values of the weighting coefficients. If some are very much smaller than others then this may indicate that a neuron is adding little to the accuracy.

A great deal has been published about the development and use of neural networks. Much of this is applicable to any form of regression. For example it is good discipline to randomly split the data into two groups – one for training and one for testing. However, failing the test will result in the engineer reconfiguring the inferential. In effect the test data is also used for training. Under these circumstances a third independent set of data should then be used for testing. This is particularly important with neural networks because they can extrapolate poorly outside the region in which they have been trained. Indeed, it should be common practice to check whether the current inputs fall within the training envelope before using the output. It would be wise to temporarily disable any control action under these circumstances until its result has been validated and the inferential re-trained as necessary.

There is a tendency to 'throw data' at neural networks. Data should be pre-processed, for example to remove outliers and to identify any cross-correlations. Consideration should be given to including derived measurements, such as ratios and cross-products. Including sufficient neurons will permit any data to be fitted, in much the same way that including additional terms in conventional regression does. However, the lack of transparency in a neural network will hide any nonsensical relationship.

It is possible to design a neural network to generate more than one output. In our example we could include additional measurements taken around the bottom of the column and also infer the $\%C_3$ in bottoms butane product. This might seem advantageous if we were including measurements that affect both product compositions – such as column pressure. However, maintenance is likely to be more difficult than supporting two separate networks.

The neural network that we have described is known as the *back propagation* (*BP*) type. Its name comes from the method used to train it and it is the one most commonly used for inferentials. There are others and it is often worth trying each of them and comparing the results.

So far we have presented neural networks as a steady-state technique. Indeed, their most common application in the process industry is inferential properties. These are usually developed on the assumption that the input data were collected at steady state. However, neural networks can also be

applied to dynamic problems. They can be set up as a dynamic model of a process, for example for use in the controllers described in Chapter 7, where historical values are used to predict future values. They can also be used as controllers, where historical values are used to determine future control moves.

14.22 Repeatability

Repeatability, as defined by the ASTM, is determined from test results which are obtained applying the same test method in the same laboratory by the same operator with the same equipment in the shortest practicable period of time using test specimens taken at random from a single quantity of homogeneous material. It can be defined as the maximum difference between two measurements that can be expected 95% of the time.

Imagine that the laboratory tests the same material twice, under these conditions, and obtains results x_1 and x_2. Both these results are taken from a normal distribution of mean μ and standard deviation σ. The average of the two results is given by

$$\bar{x} = \frac{x_1}{2} + \frac{x_2}{2} \tag{14.202}$$

Since we halve x_1 and x_2, we also halve their standard deviation. The variance of the mean will be the sum of variances.

$$\sigma_{mean}^2 = \left(\frac{\sigma}{2}\right)^2 + \left(\frac{\sigma}{2}\right)^2 \quad \text{or} \quad \sigma_{mean} = \frac{\sigma}{\sqrt{2}} \tag{14.203}$$

The 95% confidence interval is given by $1.96\sigma_{mean}$. So, if the repeatability (r) is given, the standard deviation of the test result can be derived. The approximation is that made by the ASTM.

$$\sigma = \frac{r}{1.96\sqrt{2}} \approx \frac{r}{2.8} \tag{14.204}$$

14.23 Reproducibility

The ASTM defines *reproducibility* as the expected degree of agreement between results obtained by two different laboratories employing the same procedures but different equipment when analysing splits of the same gross sample. It is determined from an *interlaboratory study* involving between 8 and 12 laboratories. It will always be greater than repeatability since it includes the variation within individual laboratories as well as the variation between them. It will typically be double the repeatability and, as with repeatability, it is determined from 2.8σ.

Normally the control engineer is more concerned with repeatability. However reproducibility becomes important, for example if a study involves laboratory results provided by a customer testing a product that is also routinely tested by the manufacturer.

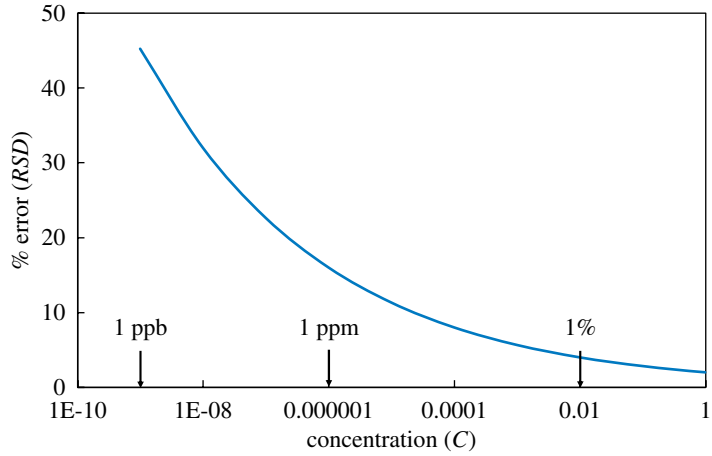

Figure 14.41 *Horwitz's curve*

For tests that measure concentration Horwitz [44,45] identified that the accuracy of a laboratory result was related to the concentration being measured. In particular, accuracy becomes very poor at very low concentrations. He studied the results of over 10,000 interlaboratory studies and developed *Horwitz's curve* as shown as Figure 14.41. The *relative standard deviation (RSD)* is defined as percentage of the measurement (C), where C is the concentration expressed as a weight fraction (or percentage).

$$RSD = 2^{\left[1-0.5\log(C)\right]} \tag{14.205}$$

From this equation

$$\log\left(RSD\right) = \log\left(2\right)\left(1-0.5\log\left(C\right)\right) \quad \text{or} \quad \log\left(\frac{RSD}{2}\right) = \log\left(C^{-0.5\log(2)}\right) \tag{14.206}$$

This gives rise to the alternative definition

$$RSD = 2C^{-0.15} \tag{14.207}$$

If we want the standard deviation (σ_H) expressed in the same units as C then the equation is further modified.

$$RSD = 100\frac{\sigma_H}{C} \qquad \text{and so} \qquad \sigma_H = \frac{RSD.C}{100} = 0.02C^{0.85} \tag{14.208}$$

Laboratory testing methods are generally accepted if their reproducibility is less than half σ_H. Although there are those who will accept a method if it performs better than $0.67\sigma_H$ and there are others who believe that the Horwitz method is unreliable.

When assessing the accuracy of a measurement made on-stream, the manufacturer's quoted reproducibility should be used rather than repeatability. For example, the repeatability of an on-stream analyser is determined by keeping the composition constant and observing how much the measurement varies. Reproducibility is determined by changing the composition and is a measure of the ability of the instrument to reproduce the measurement when a predefined set of conditions is recreated. It is therefore the more realistic parameter.

14.24 Six-Sigma

Six-sigma is a much publicised method used in quality control. However rather than, as it name suggests, it being based on six standard deviations, it is actually based on 4.5. A 1.5σ *shift* is included to reflect long-term changes thought to occur in the variation. The choice of this value is entirely empirical and often criticised by statisticians. However, it now forms part of the definition of the procedure.

From Table 14.1 we can see that the 4.5σ confidence interval is 0.9999932. Thus there is a probability 0.0000068 that the value is outside this range. Six-sigma only considers off-grade results and so uses the probability of values exceeding $\mu + 4.5\sigma$ or falling below $\mu - 4.5\sigma$, but not both. This halves the probability to 0.0000034, i.e. 3.4 defects per million.

14.25 Data Reconciliation

No process measurement can be considered perfect. The instrumentation itself is subject to error, and data collection is subject to inaccuracies in time-stamping. It is for this reason that we never expect heat and mass balances to close perfectly. However, performing such a balance is effectively a comparison between two 'opinions' of the true value – one measured directly and the other derived from the other measurements involved in the balance. Data reconciliation is a technique which uses these multiple estimates to produce an estimate which is more reliable than any of them.

Consider, as a simple example, two measurements of the same property – both subject to error. The first has a standard deviation of σ_1, the second σ_2. The values of each of these measurements can be considered to have come from two distributions with different means, i.e. μ_1 and μ_2. Our aim is to choose the most likely estimate. This will be a weighted average of the two measurements, where a and $(1-a)$ are the weighting coefficients. This estimate will therefore have a mean given by

$$\mu = a.\mu_1 + (1-a)\mu_2 \tag{14.209}$$

Provided the errors in the two measurements are not correlated, then the standard deviation will be given by

$$\sigma = \sqrt{(a.\sigma_1)^2 + ((1-a)\sigma_2)^2} = \sqrt{(\sigma_1^2 + \sigma_2^2)a^2 - 2\sigma_2^2 a + \sigma_2^2} \tag{14.210}$$

The best estimate will have the smallest standard deviation. This will occur when

$$\frac{d\sigma}{da} = 0 \quad \text{or} \quad \frac{(\sigma_1^2 + \sigma_2^2)a - \sigma_2^2}{\sqrt{(\sigma_1^2 + \sigma_2^2)a^2 - 2\sigma_2^2 a + \sigma_2^2}} = 0 \tag{14.211}$$

Thus the best choice of a is given by

$$a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{14.212}$$

Substituting this value into Equation (14.209) gives the best estimate as

$$\mu = \frac{\sigma_2^2 \mu_1 + \sigma_1^2 \mu_2}{\sigma_1^2 + \sigma_2^2} \tag{14.213}$$

And this estimate will have a standard deviation given by substituting Equation (14.212) into Equation (14.210)

$$\sigma = \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{14.214}$$

As an example, consider a distillation column which has been operating at a feed rate of 100 for the last 24 hours, as determined by a flow controller on the feed. This is the value for μ_1. The flow meter has a quoted reproducibility (1.96σ) of 4% of the instrument range which, in this case, is set at 125. Assuming no other sources of error, we can assume σ_1 has a value of 2.5.

The feed rate is also estimated from the change in level in the feed tank. Let us assume that this gives an estimate of 103. This is the value for μ_2. Tank gauging typically has a reproducibility of 2mm. Since the calculation uses the difference between two measurements, the estimate of the change in level will have a variance of $2\sigma^2$ and so the standard deviation will be 1.41mm. The effect this has on the estimate of flow depends on the magnitude of the change in tank level which, in turn, depends on the interval over which it is measured and the tank's cross-sectional area. To simplify the arithmetic let us assume that the tank level changed by 100mm. The value of σ_2 is given by $103 \times 1.41/100$ or around 1.5.

Applying Equations (14.213) and (14.214) gives a best estimate of 102.2 with a standard deviation of 1.3. This standard deviation is less than that of either measurement. Not only have we reconciled the difference between the two measurements but we have greater confidence in the resulting estimate than in either of the measurements. In other words, taking into account other measurements, even if inaccurate, can improve the accuracy of the estimate. This is illustrated as Figure 14.42.

We can add a third measurement by first replacing, in Equation (14.213), μ_2 with μ_3 and σ_2 with σ_3. We then replace μ_1 with the expression used to determine μ from the same equation and replace σ_1 with the expression used to determine σ, i.e. Equation (14.214). This gives us

$$\mu = \frac{\sigma_2^2 \sigma_3^2 \mu_1 + \sigma_1^2 \sigma_3^2 \mu_2 + \sigma_1^2 \sigma_2^2 \mu_3}{\sigma_2^2 \sigma_3^2 + \sigma_1^2 \sigma_3^2 + \sigma_1^2 \sigma_2^2} \tag{14.215}$$

Making similar substitutions in Equation (14.214) we get

$$\sigma = \frac{\sigma_1 \sigma_2 \sigma_3}{\sqrt{\sigma_2^2 \sigma_3^2 + \sigma_1^2 \sigma_3^2 + \sigma_1^2 \sigma_2^2}} \tag{14.216}$$

In practice we often have many more than two estimates of flows. In addition to those above we can derive another from the overall unit mass balance – from summing all the product flows. And we can assign a standard deviation to this measurement by considering the contribution made by the reproducibility of

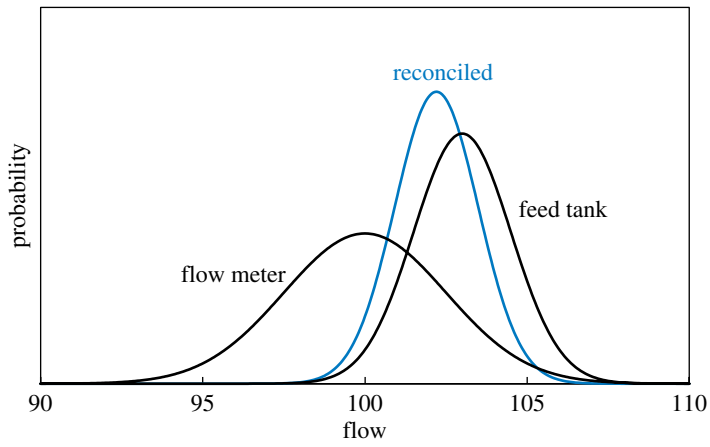

Figure 14.42 *Reconciliation of two measurements*

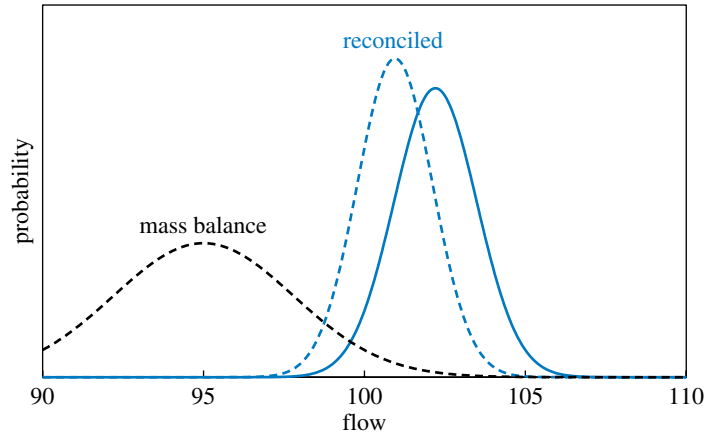

Figure 14.43 *Reconciliation of three measurements*

each instrument used in the calculation. For example our column might simply produce two products – distillate and bottoms. Let us assume that these two flows, measured over the same period, are respectively averaged at 42 and 53. Giving a total of 95, this gives us another measurement of feed rate. Like the feed meter, both product meters have a reproducibility of 4% of range. The range of the distillate meter is 90 and so the standard deviation of its measurement is 1.8. The range of the bottoms meter is 110, giving a standard deviation of 2.2. The standard deviation of the derived feed flow measurement is the square root of the sum of the squares of these two values, i.e. 2.8.

Figure 14.43 shows the effect of including this measurement. From Equation (14.215) we now find that the best estimate is 101.0. Equation (14.216) shows that the standard deviation has improved to 1.2. The difference between this best estimate and the measurement derived by mass balance is 6.0. This is more than double the standard deviation and therefore outside the quoted reproducibility, suggesting that this measurement is suspect. Whether this is caused by a problem with the distillate meter or bottoms meter

would require reconciliation of those flows using other measurements. For example we may also be able to include changes in product tank inventories or a mass balance around a downstream process. If we have analysis of the composition of the streams we can perform component balances, each of which generates another estimate. And, with sufficient temperature measurements, we might also be able to derive another value by energy balance.

Formulae using four measurements can be developed by extending Equations (14.215) and (14.216) and so on. However, for complex problems, proprietary *data reconciliation system (DRS)* software should be used. These products typically minimise the penalty function (C), where μ is the reconciled value derived from the n measurements of μ_1 to μ_n.

$$C = \sum_{i=1}^{n} \left(\frac{\mu - \mu_i}{\sigma_i} \right)^2 \tag{14.217}$$

Figure 14.44 shows the effect of applying this technique to our example process. Again it demonstrates that the best estimate is 101.

DRS software permits the user to define the complex process flowsheets and so, for example, reconcile all the measurements across a whole manufacturing site. Such information is valuable both for accounting purposes but also for process modelling. Indeed, a similar approach must be taken with the data provided to CLRTO – particularly those based on rigorous process simulation.

The software also allows the user to identify suspect measurements by considering how different the best estimate is from each measurement. In our example, the best estimate is 2.2 higher than that measured by the flow meter, i.e. within the quoted reproducibility. Had it not been then this would indicate that the meter was in need of attention. Similarly it is 0.8 lower than that calculated from tank levels – again within the expected range. Indeed, it is this systematic method of identifying, and so rectifying, measurement problems that is perhaps of the greatest benefit since it improves the quality of the raw process measurements used by all. Indeed applying DRS, without rectifying such problems, risks generating yet another set of process measurements with no increase in user confidence that they are any more reliable than the raw data.

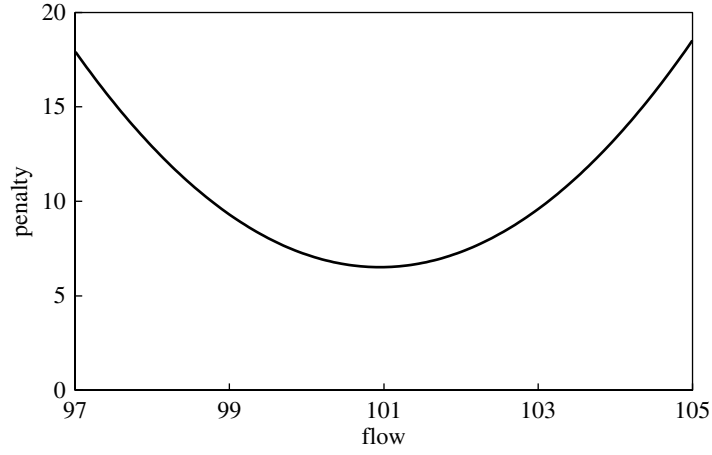

Figure 14.44 *Minimising data reconciliation penalty function*

Many DRS packages also support ‘what if?’ analysis. The inclusion of additional measurements and improving the accuracy of existing measurements can be assessed in terms of the impact they would have on overall accuracy. There are many examples where the accuracy of the estimate (derived by mass balance) of an unmeasured flow may not be significantly improved by the installation of a direct measurement. It might, however, be improved by replacing poor instrumentation elsewhere in the process. What-if analysis permits definition of the scope of the most cost effective measurement upgrade project.

The technique is not limited to the measurement of flows. It can be applied wherever there are multiple measurements of the same parameter. For example by comparing inferential (or on-stream analyser) against the laboratory result, we can decide whether the difference between the two measurements is significant enough to merit attention.

15

Mathematical Techniques

Complex mathematical techniques have created a reputation among most chemical engineers that process control is a highly theoretical subject, often beyond their understanding of the theory involved. Hopefully the reader of this book will by now have come to realise that this is far from reality. However, some of these techniques do have value. Once mastered, they can provide more convenient control design methods. Other techniques provide opportunities to improve control performance in ways which the engineer may not be aware are possible. Without them a control engineer can readily achieve a substantial improvement in controller performance. With them, perhaps a little more might be possible.

15.1 Fourier Transform

Most texts covering the Fourier transform do so in a highly mathematical way, making it difficult for the control engineer to identify where its application might be beneficial. Here we restrict its use to identifying cyclic disturbances to process measurements. Often such disturbances are not immediately obvious, presenting themselves as random noise. This is a common problem with averaging level control. Applying this technique has identified many controllers that appear to be working well, using the available surge capacity, but the variation in level is in fact a very slow oscillation (disguised by process disturbances) caused by excessive integral action. The technique can also help diagnose control valve problems, such as stiction and hysteresis.

Since control engineers deal largely with process data collected at a fixed time interval we will focus on the *discrete Fourier transform (DFT)*. Strictly it is the *real DFT* where the input is restricted to real data. There is also an *imaginary DFT* but this has no application here.

Fourier showed that any signal can be decomposed into a number of sinusoidal signals. Each of these signals will have a different *frequency* and *amplitude*; further they will not necessarily be *in phase*. We need to be careful with the definitions and units of measurement of these terms. If we consider the continuous function for the k^{th} frequency

$$x_k = a.\sin\left(2\pi f_k t + \phi\right) \tag{15.1}$$

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

The term a is known as the *peak amplitude*. Many engineers will use the term 'amplitude' to describe the *peak-to-peak* distance given by $2a$. Amplitude can also be referred to as *energy* or *power*. If time (t) is measured in seconds, then the frequency (f_k) will be in Hz (sec^{-1}). The factor 2π is required to convert f_k from cycles per unit time to radians per unit time. Similarly the phase shift (ϕ) is expressed in radians.

With discrete Fourier transforms the frequency (k) is expressed as the number of cycles that occur within the period over which the input data are collected. The term $f_k t$ is replaced by ki/N. The sine wave at this frequency has a peak amplitude of a_k and now comprises a series of discrete values of x_k. N is the total number of values collected in the period and i is the index (ranging from 0 to N-1) of each value of x_k.

$$\left(x_k\right)_i = a_k.\sin\left(\frac{2\pi ki}{N} + \phi_k\right) \tag{15.2}$$

This equation can be rewritten as

$$\left(x_k\right)_i = a_k\left[\sin\left(\phi_k\right)\cos\left(\frac{2\pi ki}{N}\right) + \cos\left(\phi_k\right)\sin\left(\frac{2\pi ki}{N}\right)\right] \tag{15.3}$$

The result of the Fourier transform is, for each frequency, a combination of in-phase sine and cosine waves of peak amplitude p_k and q_k respectively. For example, at frequency k, the series of x_k values is given by

$$\left(x_k\right)_i = p_k.\cos\left(\frac{2\pi ki}{N}\right) + q_k.\sin\left(\frac{2\pi ki}{N}\right) \tag{15.4}$$

Comparing this with Equation (15.3) gives

$$p_k = a_k.\sin\left(\phi_k\right) \qquad \text{and} \qquad q_k = a_k.\cos\left(\phi_k\right) \tag{15.5}$$

The peak amplitude is therefore given by

$$a_k = \sqrt{p_k^2 + q_k^2} \tag{15.6}$$

One might reasonably deduce that the phase angle is given by

$$\phi_k = \tan^{-1}\left(\frac{p_k}{q_k}\right) \tag{15.7}$$

However, this will only generate a value for ϕ_k in the range $-\pi/2$ to $\pi/2$. Further it is indeterminate when q_k is zero. To obtain values from the full range of $-\pi$ to π requires the use of the *arctangent2 (atan2)* function. This is shown in Figure 15.1 and is defined as follows:

$$\text{atan2}\left(\frac{p_k}{q_k}\right) = \tan^{-1}\left(\frac{p_k}{q_k}\right) \qquad \text{when} \quad q_k > 0 \tag{15.8}$$

$$\text{atan2}\left(\frac{p_k}{q_k}\right) = \tan^{-1}\left(\frac{p_k}{q_k}\right) + \pi \qquad \text{when} \quad p_k \geq 0 \quad \text{and} \quad q_k < 0 \tag{15.9}$$

$$\text{atan2}\left(\frac{p_k}{q_k}\right) = \tan^{-1}\left(\frac{p_k}{q_k}\right) - \pi \qquad \text{when} \quad p_k < 0 \quad \text{and} \quad q_k < 0 \tag{15.10}$$

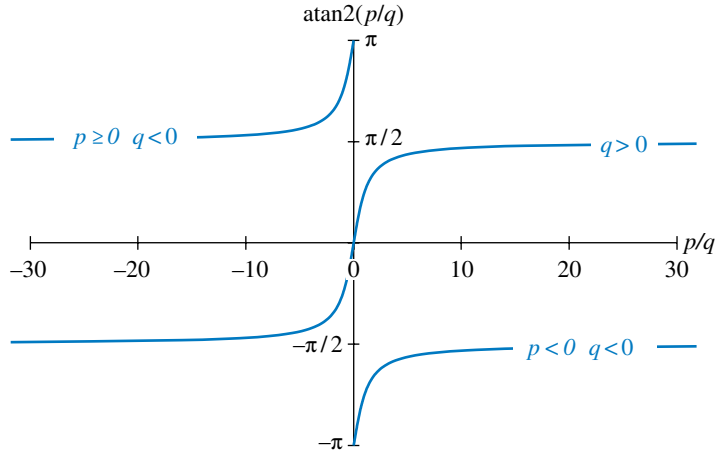

Figure 15.1 *Arctangent2 function*

$$\text{atan2}\left(\frac{p_k}{q_k}\right) = \frac{\pi}{2} \qquad \text{when} \quad p_k \geq 0 \quad \text{and} \quad q_k = 0 \tag{15.11}$$

$$\text{atan2}\left(\frac{p_k}{q_k}\right) = -\frac{\pi}{2} \qquad \text{when} \quad p_k < 0 \quad \text{and} \quad q_k = 0 \tag{15.12}$$

We need to consider the conditions under which the transformation can be applied. Firstly, the data must be a repeating waveform. If an insufficient number of values are analysed, then the resulting distortion is known as *leakage*. Secondly, if an analog signal contains no frequency higher than or equal to f_{max}, it can be reconstructed from the digital signal providing the sampling exceeds $2f_{max}$ samples per second. This is known as the *Nyquist-Shannon sampling theorem*. For example, many process historians collect data at one minute intervals; in which case the maximum frequency detectable will be 30 hr^{-1}, i.e. 0.5 min^{-1} or 1/120 Hz. Collecting data at too low a frequency causes distortion known as *aliasing*. The maximum value of k is therefore $N/2$; k has a minimum value of zero. Thus the Fourier transform will comprise $N/2+1$ sine waves and $N/2+1$ cosine waves. It should be noted therefore that N should be an even number and that only discrete frequencies that are an exact multiple of the data collection interval are included. Some texts suggest that N should be a power of 2 (for example 512, 1024 …). This permits the use of the *fast Fourier transform* technique, substantially reducing the calculations required and hence the execution time of any analysis tool. However, this only becomes an issue when analysing very high frequencies and is of less advantage for process data.

The coefficients (p) for each frequency (k) are determined from the actual sampled data (x) using the formulae

$$p_k = \frac{2}{N}\sum_{i=0}^{N-1} x_i \cos\left(\frac{2\pi ki}{N}\right) \tag{15.13}$$

Except

$$p_0 = \frac{1}{N}\sum_{i=0}^{N-1} x_i \qquad \text{and} \qquad p_{N/2} = \frac{1}{N}\sum_{i=0}^{N-1} x_i \cos(2\pi i) \tag{15.14}$$

The reason that the calculations for p_0 and $p_{N/2}$ are slightly different has to do with *bandwidth*. Bandwidth is the difference between the highest and lowest frequencies in the band. The total bandwidth is split into $N/2$ bands, but there are $N/2+1$ frequencies. The width of each band, for the frequencies from $k = 1$ to $k = N/2-1$, is $2/N$. However the frequencies at each end, i.e. $k = 0$ and $k = N/2$, have a bandwidth of half this value, or $1/N$.

The coefficient p_0 is applied to the cosine of zero, which is unity. It is therefore the mean of all the values collected and represents the *offset* of the signal from zero.

The coefficients (q) for each frequency (k) are also determined from the actual sampled data (x) but using the formulae

$$q_k = \frac{2}{N}\sum_{i=0}^{N-1} x_i \sin\left(\frac{2\pi ki}{N}\right) \tag{15.15}$$

Since q_0 is derived from the sine of zero, it will also be zero. Similarly $q_{N/2}$ will be zero since is it derived from the sine of integer multiples of π, each of which will be zero.

The analog signal can be reconstructed from the *inverse discrete Fourier transform (IDFT)*.

$$x_i = \sum_{k=0}^{N/2} p_k \cos\left(\frac{2\pi ki}{N}\right) + \sum_{k=0}^{N/2} q_k \sin\left(\frac{2\pi ki}{N}\right) \tag{15.16}$$

Figure 15.2 shows 60 values collected at one minute intervals for an hour – typically what might be readily available from a process historian. The coefficients p_0 to p_{30} were calculated using Equations (15.13) and (15.14); q_0 to q_{30} using Equation (15.15). These were then used in Equation (15.16) to reconstruct the analog signal shown as the coloured line. This simply confirms that the method works and is a useful check that the calculations have been performed correctly. It should be noted that for digital signals the reconstructed analog signal will pass through all the discrete values; it is not a curve of best fit but an exact solution. Neglecting q_0 and $q_{N/2}$, which are both zero, we have calculated N coefficients to fit a curve through N points.

The value of the technique is that, by applying Equation (15.6), we can determine the peak amplitude of every frequency and present this as the *power spectrum* shown as Figure 15.3. We have converted a

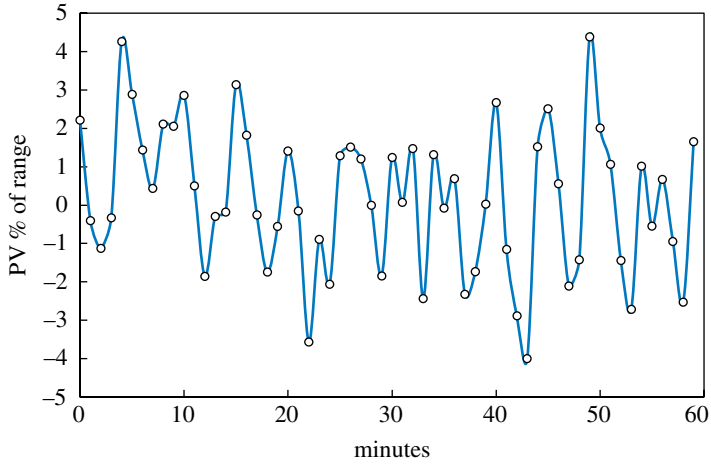

Figure 15.2 *Apparently noisy measurement*

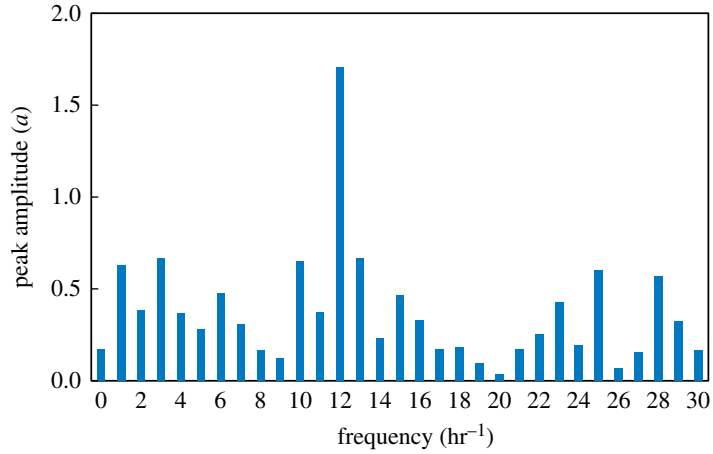

Figure 15.3 *Power (or frequency) spectrum*

signal that is in the *time domain* to one in the *frequency domain*. In our example the average signal is approximately zero and so a_0 is small. If this is not the case, then the zero frequency should be omitted from the chart so that it does not dwarf the others. Alternatively the mean (μ) can be subtracted from each value of x before calculating the amplitudes. This affects only p_0, reducing it (and hence a_0) to zero. Figure 15.4 shows the same power spectrum but based on the period (wavelength) of oscillation.

The power spectrum shows that the apparently random noise in the process measurement is dominated by a waveform that completes 12 cycles within the period covered by the data, i.e. it has a period of 5 minutes. This might give some clue to the source of the oscillation – particularly if the technique is also applied to values collected over the same period from potential sources. The peak amplitude at this frequency (a_{12}) is about 1.7, giving a peak-to-peak amplitude of 3.4.

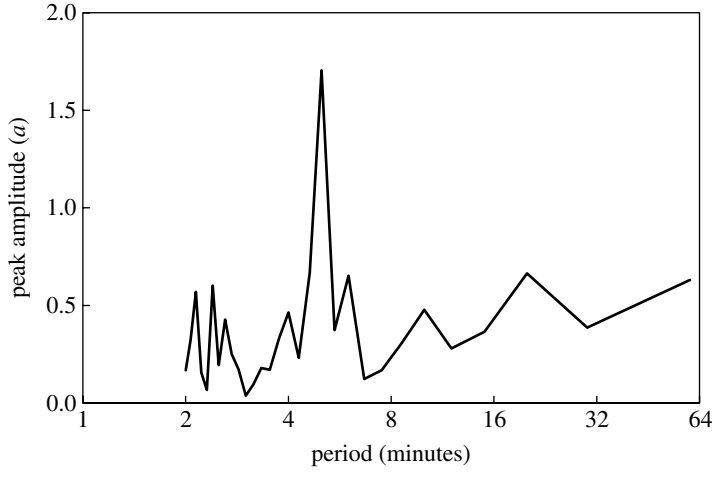

Figure 15.4 *Period spectrum*

If required, we can also obtain the *phase spectrum* by applying the appropriate choice from Equations (15.8) to (15.12). This is shown in Figure 15.5 and gives a value for ϕ_{12} of 1.63 radians.

Using the values a_{12} and ϕ_{12} in Equation (15.1) we can superimpose the waveform on the original signal, as shown in Figure 15.6. There is a strong correlation between the two curves with a value for Pearson R of 0.64. While not accounting for all the noise, it does explain a large part of the variation. Very similar values for a_{12} and ϕ_{12} would be obtained by regressing the coefficients to give a wave of best fit.

A limitation of this discrete method is that the frequency spectrum can only contain frequencies equivalent to integer values of k. So, for example, if we increase the number of values collected from 60 to 62 the wavelength corresponding to k having a value of 12 will increase to 5.17 minutes (62/12). If there is truly dominant waveform, with a wavelength of 5 minutes, its dominance will be masked by adjacent frequencies. Figure 15.7 shows how the spectrum will now appear. Superimposing the dominant waveform on the

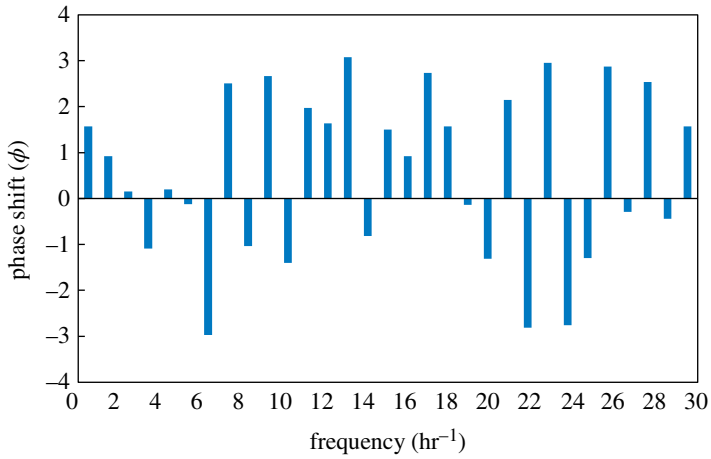

Figure 15.5 *Phase spectrum*

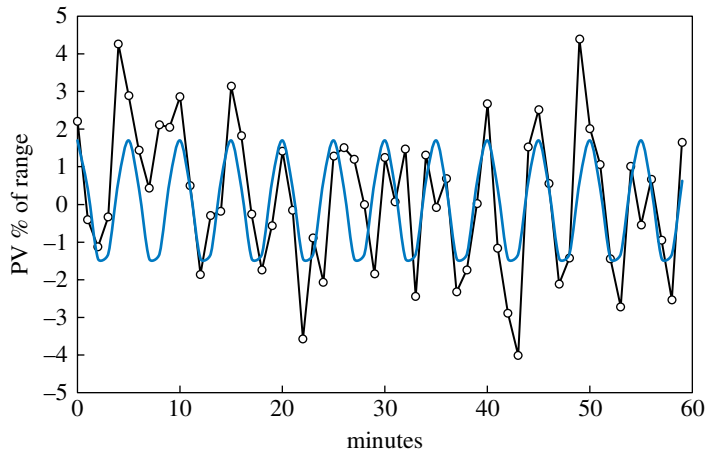

Figure 15.6 *Superimposing dominant wave form on original measurement*

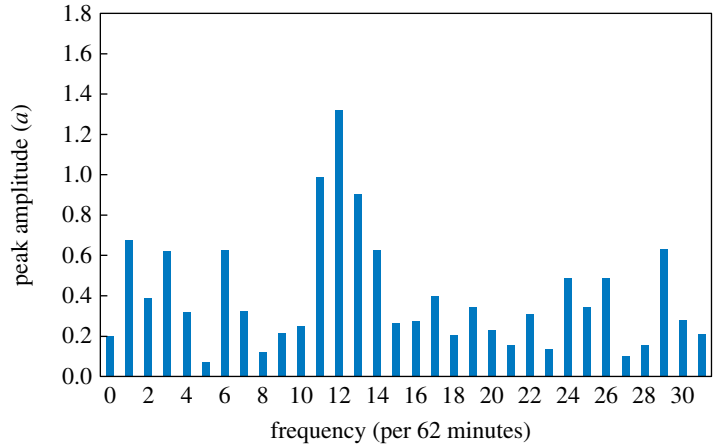

Figure 15.7 *Frequency spectrum derived with two additional data points*

original signal will show a poorer match; indeed, in this case Pearson R reduces to 0.48. As might be expected, synthesising a signal by also including the two frequencies either side of the dominant one increases R to 0.67 – close to that derived from the original set of data. It is therefore good practice to analyse different sets of data, making small changes to the number of values included in each set.

While DFT will highlight a problem with a controller, it does not necessarily mean that it is caused by the controller. The problem may persist even if the controller is switched to manual – so demonstrating that the cause may be elsewhere in the process. To illustrate this, Figure 15.8 shows data collected from another controller (PV_2) superimposed on the trend for the controller analysed previously (PV_1). It would be difficult to conclude from these trends whether changes in PV_2 are causing the variation in PV_1. Indeed, the scatter plot shown as Figure 15.9 shows no obvious relationship. However, we can apply DFT to generate the power spectrum for PV_2 and plot the amplitude of each frequency against the corresponding amplitude

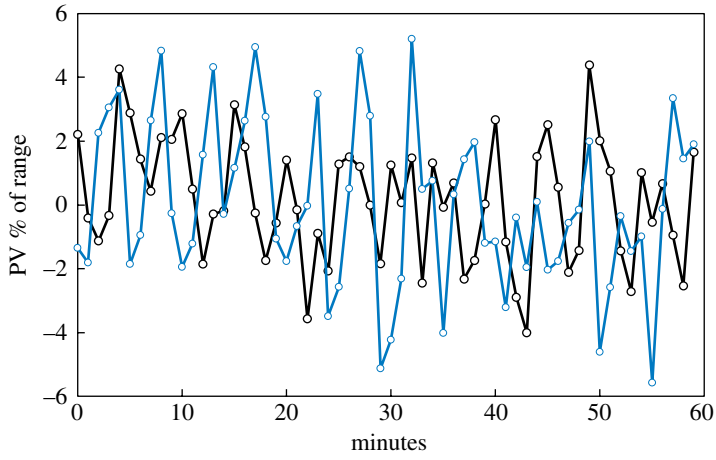

Figure 15.8 *Addition of second noisy measurement*

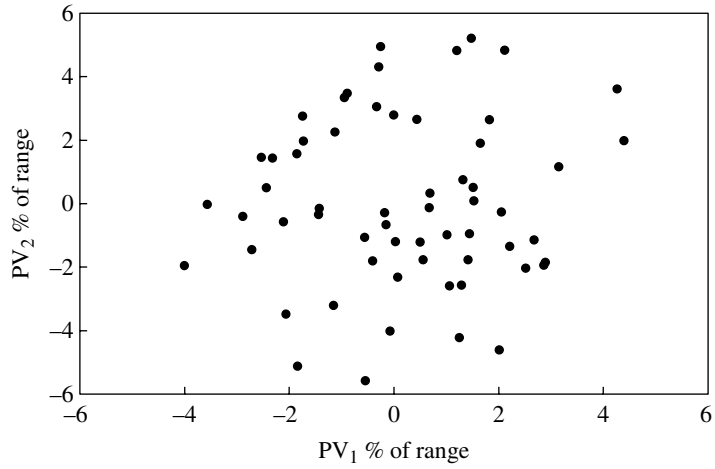

Figure 15.9 *No apparent correlation between measurements*

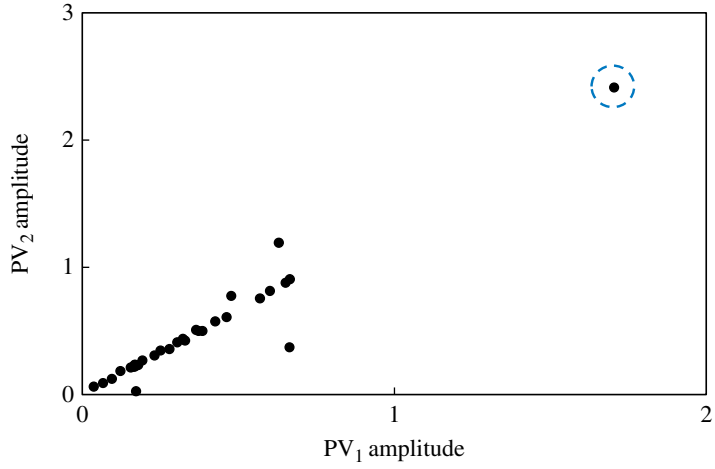

Figure 15.10 *Correlation between frequency spectra*

from the PV_1 power spectrum. Figure 15.10 shows not only are both process values dominated by the frequency of 12 cycles/minute (highlighted) but that they contain many other frequencies in common. The very strong correlation (Pearson R^2 of 0.92) suggests that oscillation in PV_2 may be the cause of that in PV_1 – or vice versa. This technique is used by several controller diagnostics products. Indeed, these will often analyse all the process values and generate a *power spectrum correlation map* – a grid using colour to show where correlations are strong. While valuable, this method does not distinguish cause from effect. To confirm the cause, it is still necessary to switch to manual any controller thought to be the source of the problem.

15.2 Recursive Filters

We covered in Chapter 5 a range of techniques for filtering noise. One of these, the first order exponential filter, is an example of a *recursive filter*. This simply means that it uses the previous output (Y) from the filter, in addition to the input (X), to determine the current output. In general the filter will have the following form:

$$Y_n = a_0 X_n + a_1 X_{n-1} + a_2 X_{n-2} \ \ldots + b_1 Y_{n-1} + b_2 Y_{n-2} \ \ldots . \tag{15.17}$$

Provided there is at least one b coefficient, the filter is recursive. In the case of the first order filter a_0 is P and b_1 is 1-P. All the other coefficients are zero.

Higher order filters will achieve the same level of noise reduction but with less signal distortion of the base signal. For example a second order exponential filter, with two different lags, can be developed by connecting two first order filters in series. The first would be as Equation (5.29) generating an intermediate output Y^* that becomes the input to a second filter, i.e.

$$Y_n^* = P_1 Y_{n-1}^* + \left(1 - P_1\right) X_n \tag{15.18}$$

$$Y_n = P_2 Y_{n-1} + \left(1 - P_2\right) Y_n^* \tag{15.19}$$

Substituting Equation (15.18) into Equation (15.19)

$$Y_n = P_2 Y_{n-1} + \left(1 - P_2\right)\left(P_1 Y_{n-1}^* + \left(1 - P_1\right) X_n\right) \tag{15.20}$$

Writing Equation (15.19) for the preceding scan

$$Y_{n-1} = P_2 Y_{n-2} + \left(1 - P_2\right) Y_{n-1}^* \qquad \text{or} \qquad Y_{n-1}^* = \frac{Y_{n-1} - P_2 Y_{n-2}}{1 - P_2} \tag{15.21}$$

Substituting into Equation (15.20)

$$Y_n = \left(P_1 + P_2\right) Y_{n-1} - P_1 P_2 Y_{n-2} + \left(1 - P_1\right)\left(1 - P_2\right) X_n \tag{15.22}$$

We can increase the number of lags in the filter to three by repeating the exercise above. Equation (15.22) will generate the intermediate output Y^* that now becomes the input to a third filter.

$$Y_n^* = \left(P_1 + P_2\right) Y_{n-1}^* - P_1 P_2 Y_{n-2}^* + \left(1 - P_1\right)\left(1 - P_2\right) X_n \tag{15.23}$$

$$Y_n = P_3 Y_{n-1} + \left(1 - P_3\right) Y_n^* \tag{15.24}$$

Substituting Equation (15.23) into Equation (15.24)

$$Y_n = P_3 Y_{n-1} + \left(1 - P_3\right)\left(\left(P_1 + P_2\right) Y_{n-1}^* - P_1 P_2 Y_{n-2}^* + \left(1 - P_1\right)\left(1 - P_2\right) X_n\right) \tag{15.25}$$

Writing Equation (15.24) for the preceding two scans

$$Y_{n-1} = P_3 Y_{n-2} + \left(1 - P_3\right) Y_{n-1}^* \qquad \text{hence} \qquad Y_{n-1}^* = \frac{Y_{n-1} - P_3 Y_{n-2}}{1 - P_3} \tag{15.26}$$

$$Y_{n-2} = P_3 Y_{n-3} + (1-P_3) Y^*_{n-2} \qquad \text{hence} \qquad Y^*_{n-2} = \frac{Y_{n-2} - P_3 Y_{n-3}}{1-P_3} \tag{15.27}$$

Substituting Equations (15.26) and (15.27) into Equation (15.25)

$$Y_n = (P_1+P_2+P_3) Y_{n-1} - (P_1P_2 + P_1P_3 + P_2P_3) Y_{n-2} + P_1P_2P_3 Y_{n-3} + (1-P_1)(1-P_2)(1-P_3) X_n \tag{15.28}$$

Higher order filters of this sort are known as *Butterworth filters*. For an n^{th} order filter, if we assume we have n equal filter lags, the overall lag is given by the modified form of Equation (5.39).

$$\tau_f = \frac{\sqrt{(K_f)^{-2/n} - 1}}{2\pi f} \tag{15.29}$$

Assuming attenuation (K_f) of 0.1 is required for noise of frequency (f) of 12 min^{-1}, Figure 15.11 shows the effect on a step input of combining n such lags. The overall lag, as defined by the 63.2% response, is typically halved by changing from a first order to a third order filter. In situations where a first order filter exceeds the process lag by more 20% but by less than 40%, a second or third order filter might be considered. Table 15.1 shows the coefficients required for higher order filters – assuming all the P coefficients are set to the same value. Note that the coefficients always sum to unity. In each case the value of P is chosen between 0 and 1 to specify the severity of the filter. Further improvement is possible by using unequal values. But their selection requires detailed knowledge of the frequency distribution of the noise and real changes in the base signal. The difference in performance would go unnoticed on a real process and the effort required in installing and tuning them is not justified. The approach is too sophisticated for the process industry and certainly beyond the scope of this book.

If we replace X with MV and Y with PV then Equation (15.22) can be written in the form

$$PV_n = a_1 PV_{n-1} + a_2 PV_{n-2} + b_1 MV_{n-\theta/ts} \tag{15.30}$$

$$a_1 = \exp\left(\frac{-ts}{\tau_1}\right) + \exp\left(\frac{-ts}{\tau_2}\right) \qquad a_2 = -\exp\left(\frac{-(\tau_1+\tau_2)ts}{\tau_1\tau_2}\right) \tag{15.31}$$

$$b_1 = (1-e^{-ts/\tau_1})(1-e^{-ts/\tau_2}) \tag{15.32}$$

Equation (15.30) has a form similar to the second order process model described by Equation (2.30) and indeed can be used to identify the process dynamics of a second order process.

Control engineers are normally concerned with *low pass* filters, i.e. those which remove high frequency noise but allow low frequency process changes to pass through. However, it is equally possible to choose coefficients to produce a *high pass* filter. The simplest would have the form

$$Y_n = \left(1 - \frac{P}{2}\right) X_n + \left(\frac{P}{2} - 1\right) X_{n-1} + (1-P) Y_{n-1} \tag{15.33}$$

Figure 15.12 shows a test signal formed by combining a low frequency and a high frequency sine wave. Figure 15.13 compares the outputs of the two types of filter. As expected the low pass filter largely removes the high frequency element and allows the low frequency changes to pass through, albeit slightly distorted. The high pass filter does the opposite.

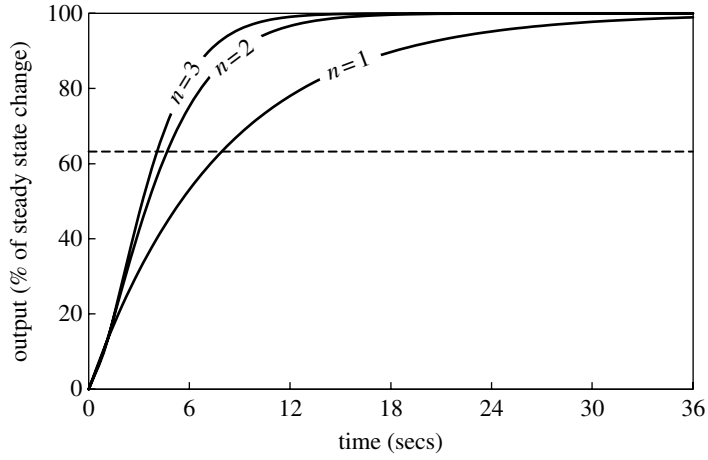

Figure 15.11 *Reduction in lag from using higher order filters*

Table 15.1 *Coefficients for exponential filters*

Coefficient	1st order	2nd order	3rd order	4th order	5th order
a_0	$1-P$	$(1-P)^2$	$(1-P)^3$	$(1-P)^4$	$(1-P)^5$
b_1	P	$2P$	$3P$	$4P$	$5P$
b_2		$-P^2$	$-3P^2$	$-6P^2$	$-10P^2$
b_3			P^3	$4P^3$	$10P^3$
b_4				$-P^4$	$-5P^4$
b_5					P^5

It is also possible to produce a *band pass* filter [46], where only frequencies falling within a defined bandwidth are allowed through. This form of filter is used in controller performance monitoring techniques to pre-process data before analysis to check for oscillation caused by poor controller tuning or a control valve problem. Oscillation is detected by checking whether there is a regular interval between the controller error changing sign (zero crossing). To remove crossings not caused by oscillation, the filter bandwidth is set to remove both low frequencies (associated with genuine process disturbances) and high frequencies (associated with measurement noise).

To design the filter, we first choose the centre frequency (f) and the bandwidth (w) that we wish the filter to pass. Both these parameters are expressed as cycles per sample interval. The coefficients for use in Equation (15.17) are then determined from the following equations.

$$R = 1 - 3w \tag{15.34}$$

$$K = \frac{1 - 2R\cos(2\pi f) + R^2}{2 - 2\cos(2\pi f)} \tag{15.35}$$

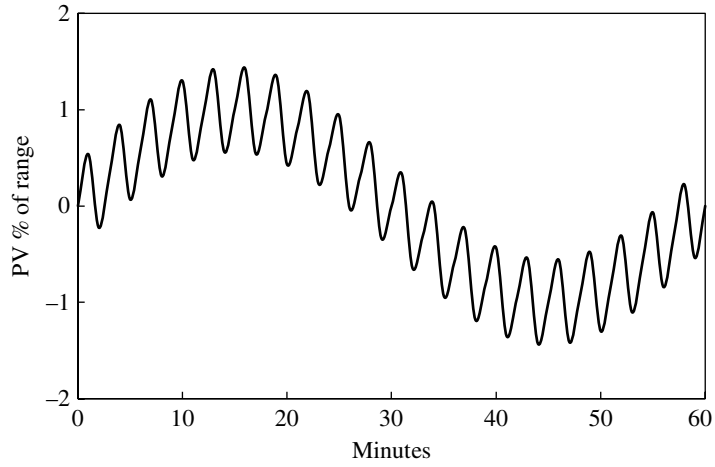

Figure 15.12 *Test signal for filters*

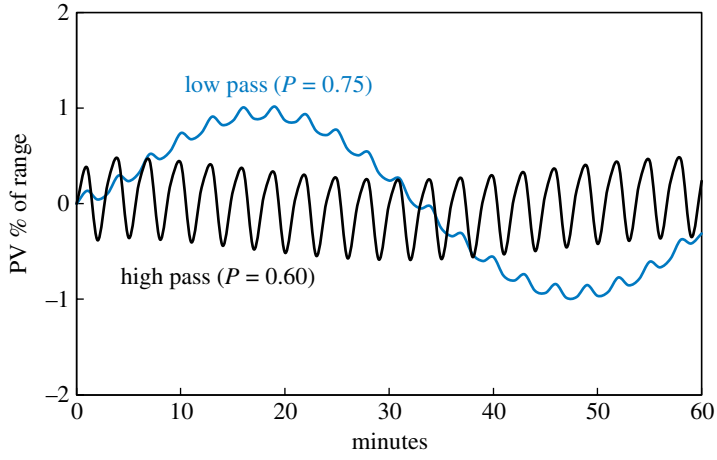

Figure 15.13 *Performance of low pass and high pass filters*

$$a_0 = 1 - K \tag{15.36}$$

$$a_1 = 2(K - R)\cos(2\pi f) \tag{15.37}$$

$$a_2 = R^2 - K \tag{15.38}$$

$$b_1 = 2R\cos(2\pi f) \tag{15.39}$$

$$b_2 = -R^2 \tag{15.40}$$

To illustrate its effectiveness we have used the noisy signal shown in Figure 15.2. From the Fourier analysis completed in Section 15.1 we know that this is dominated by a waveform that completes 12 cycles

within the one hour data collection period. The sample interval is 1 minute and so we choose 0.2 cycles per minute as f. We choose 1/60 cycles per minute as w, corresponding to the bandwidth we used in the Fourier analysis. Figure 15.14 shows the effect the filter has on the power spectrum and Figure 15.15 shows the resulting measurement.

A *band stop* (also known as a *notch* if the *stop-band* is narrow) filter can also be designed to instead remove frequencies falling within the defined bandwidth. This has the same form except that the a coefficients are defined differently.

$$a_0 = K \tag{15.41}$$

$$a_1 = -2K\cos(2\pi f) \tag{15.42}$$

$$a_2 = K \tag{15.43}$$

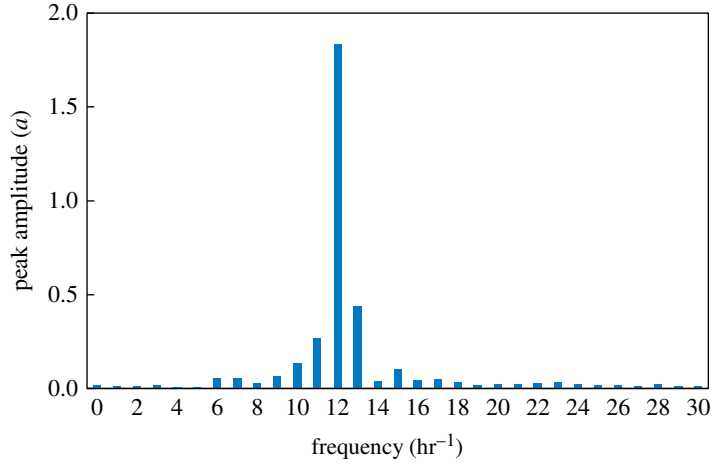

Figure 15.14 *Effect of band pass filter on frequency spectrum*

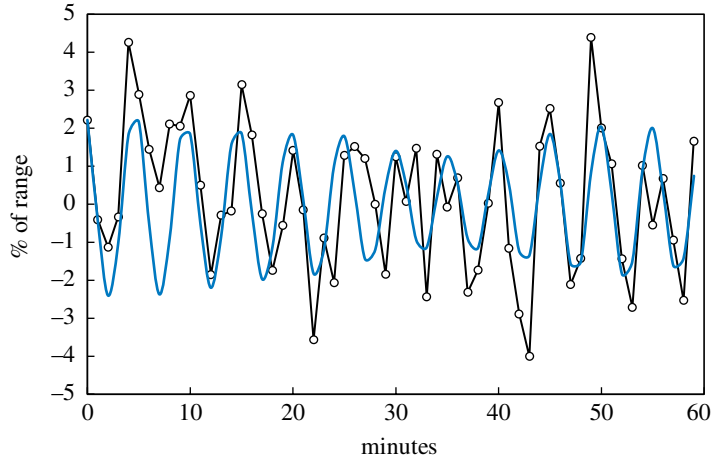

Figure 15.15 *Effect of band pass filter on process measurement*

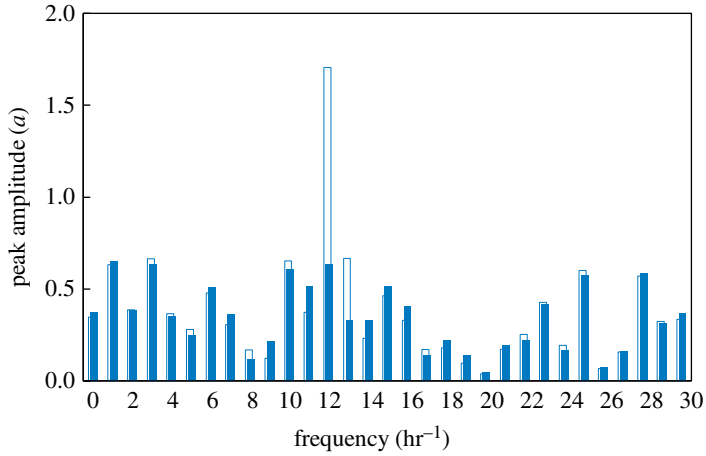

Figure 15.16 *Effect of band stop filter on frequency spectrum*

Using the same noisy signal as for the band pass filter, with the same values of f and w, Figure 15.16 shows the effect of a band stop filter. Comparing this to Figure 15.3, we can see that the amplitude of the dominant frequency has been substantially reduced with little effect on the other frequencies. If we cannot resolve the problem at source, then such a filter might be an effective alternative solution.

15.3 Lagrangian Interpolation

Lagrangian interpolation is a technique for determining intermediate values on curves defined by a series of data points. For example we may require an adaptive controller where tuning constants must be changed as process conditions vary. We have tuning known to work at several different conditions but need to interpolate between these sets of tuning constants as the conditions change. Similarly distillation curves, as we saw in Figure 12.13, are recorded as a series of temperatures below which (typically) 0, 5, 10, 20, 30, 50, 70, 80, 90, 95 and 100% of the liquid evaporates. Any missing values can be interpolated. But it also sometimes necessary to determine the percentage evaporated at a defined temperature. This temperature is unlikely to be one of those recorded. By transposing the percentage and temperature data, the required value may be derived from interpolation.

Supposing we are given the following data points

$$(x_1, y_1) = (0,1) \tag{15.44}$$

$$(x_2, y_2) = (1,2) \tag{15.45}$$

$$(x_3, y_3) = (2,7) \tag{15.46}$$

$$(x_4, y_4) = (3,22) \tag{15.47}$$

$$(x_5, y_5) = (5,106) \tag{15.48}$$

$$(x_6, y_6) = (6,187) \tag{15.49}$$

$$(x_7, y_7) = (7,220) \tag{15.50}$$

$$(x_8, y_8) = (8,230) \tag{15.51}$$

We wish to determine the value of y corresponding to $x = 4$. We could linearly interpolate between (x_4, y_4) and (x_5, y_5) as follows.

$$y = y_4 + \frac{x - x_4}{x_5 - x_4}(y_5 - y_4) \tag{15.52}$$

This would give us a value of 64, but the equation can be re-written as

$$y = \frac{y_5 - y_4}{x_5 - x_4}x + \frac{x_5 y_4 - x_4 y_5}{x_5 - x_4} \tag{15.53}$$

This is the equation of the straight line connecting (x_4, y_4) and (x_5, y_5) with the slope (m) and y-intercept (c) given by

$$m = \frac{y_5 - y_4}{x_5 - x_4} \tag{15.54}$$

$$c = \frac{x_5 y_4 - x_4 y_5}{x_5 - x_4} \tag{15.55}$$

It can be similarly be rewritten as a weighted average of y_4 and y_5.

$$y = \frac{x_5 - x}{x_5 - x_4}y_4 + \frac{x - x_4}{x_5 - x_4}y_5 \tag{15.56}$$

And this equation can be rearranged to

$$y = \frac{x - x_5}{x_4 - x_5}y_4 + \frac{x - x_4}{x_5 - x_4}y_5 \tag{15.57}$$

We shall see that this is identical to the first order Lagrangian interpolation. In general we have an ordered series of m points, (u_1, v_1) to (u_m, v_m), on a curve. We wish to calculate the interpolated value (v) corresponding to a chosen intermediate value (u). For interpolation of order n, this is derived from the sum of the products (P), i.e.

$$v = \sum_{i=1}^{n} P_i \tag{15.58}$$

Each product is determined from:

$$P_i = v_j \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{u - u_j}{u_i - u_j} \tag{15.59}$$

(The reader may be unfamiliar with the Π operator which multiplies a series of values. It is similar to the more common Σ operator – which sums a series.)

Combining Equations (15.58) and (15.59) and expanding gives

$$\begin{aligned} v = & \frac{(u-u_2)(u-u_3)....(u-u_n)}{(u_1-u_2)(u_1-u_3)....(u_1-u_n)} v_1 \\ & + \frac{(u-u_1)(u-u_3)....(u-u_n)}{(u_2-u_1)(u_2-u_3)....(u_2-u_n)} v_2 \\ & + ... \\ & + \frac{(u-u_1)(u-u_2)....(u-u_{n-1})}{(u_n-u_1)(u_n-u_2)....(u_n-u_{n-1})} v_n \end{aligned} \tag{15.60}$$

So a first order interpolation would be given by

$$v = \frac{(u-u_2)}{(u_1-u_2)} v_1 \quad + \frac{(u-u_1)}{(u_2-u_1)} v_2 \tag{15.61}$$

In our example u is x, u_1 is x_4 and u_2 is x_5. Similarly v is y, v_1 is y_4 and v_2 is y_5. Equation (15.61) is thus identical to Equation (15.57).

A second order interpolation would be given by

$$v = \frac{(u-u_2)(u-u_3)}{(u_1-u_2)(u_1-u_3)} v_1 \quad + \frac{(u-u_1)(u-u_3)}{(u_2-u_1)(u_2-u_3)} v_2 + \frac{(u-u_1)(u-u_2)}{(u_3-u_1)(u_3-u_2)} v_3 \tag{15.62}$$

Because this involves an odd number of data points we cannot place our interpolated point midway. Since (u, v) is $(4, y)$, choosing (x_3, y_3) as (u_1, v_1), (x_4, y_4) as (u_2, v_2) and (x_5, y_5) as (u_3, v_3) gives

$$y = \frac{(4-3)(4-5)}{(2-3)(2-5)} \times 7 \quad + \frac{(4-2)(4-5)}{(3-2)(3-5)} \times 22 + \frac{(4-2)(4-3)}{(5-2)(5-3)} \times 106 = 55 \tag{15.63}$$

Alternatively choosing (x_4, y_4) as (u_1, v_1), (x_5, y_5) as (u_2, v_2) and (x_6, y_6) as (u_3, v_3) gives

$$y = \frac{(4-5)(4-6)}{(3-5)(3-6)} \times 22 \quad + \frac{(4-3)(4-6)}{(5-3)(5-6)} \times 106 + \frac{(4-3)(4-5)}{(6-3)(6-5)} \times 187 = 51 \tag{15.64}$$

These two cases are shown on Figure 15.17. In each case we have fitted a second order function to three points. The actual functions can be derived from Equations (15.63) and (15.64) by replacing u with x rather than the value 4. The curves show the slight difference in the estimate of y when x is 4.

Of course, all four points could be used in a third order interpolation.

$$\begin{aligned} y = & \frac{(4-3)(4-5)(4-6)}{(2-3)(2-5)(2-6)} \times 7 \quad + \frac{(4-2)(4-5)(4-6)}{(3-2)(3-5)(3-6)} \times 22 \\ & + \frac{(4-2)(4-3)(4-6)}{(5-2)(5-3)(5-6)} \times 106 + \frac{(4-2)(4-3)(4-5)}{(6-2)(6-3)(6-5)} \times 187 = 53 \end{aligned} \tag{15.65}$$

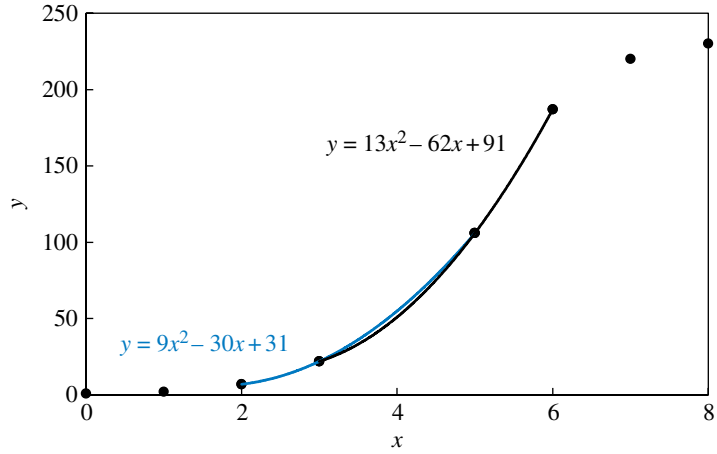

Figure 15.17 *Second order Lagrangian interpolation*

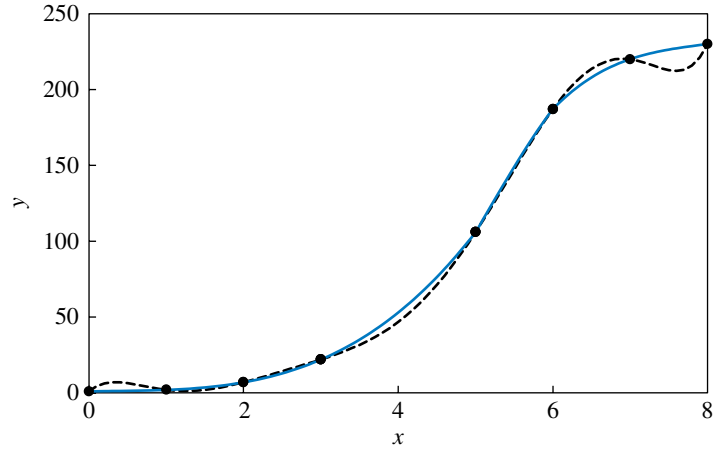

Figure 15.18 *Overfitting Lagrangian interpolation*

The maximum order possible is one less than the number of data points. In our example we could therefore choose up to a seventh order interpolation, however, too high a choice may result in over-fitting the data. This is illustrated in Figure 15.18 where the third and seventh order interpolations have been plotted (as the coloured and dashed lines respectively) for all values of x in the range 0 to 8. Although increasing the order to 7 has little effect on the value interpolated for $x = 4$, it would cause inaccuracies for other values – most notably for $x = 7.6$.

Lagrangian interpolation, like linear interpolation, is possible outside the range of data points but the result will be increasingly unreliable the further we move away from the range.

Although similar in appearance, deriving an equation using Lagrangian interpolation is different from deriving one by regression analysis. The first will necessarily pass through all the points used in the calculation. Regression will give the same result if the order of the equation regressed is one less than the

number of points. However, if more points are added, these are ignored by Lagrangian interpolation unless the order is also increased. Regression will take account of the additional points by deriving a curve of best fit – that may indeed pass through none of the points.

15.4 Padé Approximation

The Padé approximation is used extensively in control theory to convert an exponential term into a polynomial function. It is derived from the *Maclaurin series*, which in turn is derived from the *Taylor series* which sums successive derivatives of the function f evaluated at a chosen value a.

$$f(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2!} + f'''(a)\frac{(x-a)^3}{3!} \ldots . \tag{15.66}$$

By setting a to zero we obtain the Maclaurin series

$$f(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2!} + f'''(0)\frac{x^3}{3!} \ldots . \tag{15.67}$$

If $f(x)$ is e^x, then all the derivatives are also e^x and we obtain the series

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} \ldots \tag{15.68}$$

This will only converge for values of x between -1 and $+1$.

We want to keep the polynomial as simple as possible but truncating the series to a first order function, as Figure 15.19 shows, gives a very poor approximation. We can see that we need to go to third order before the approximation is reliable. A common requirement is an approximation for the Laplace transform for deadtime $e^{-\theta s}$. From Equation (15.68), the first order approximation would be

$$e^{-\theta s} = 1 - \theta s \tag{15.69}$$

Rearranging Equation (15.68) gives an alternative approach.

$$e^{-x} = \frac{1}{1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} \ldots} . \tag{15.70}$$

This gives a more accurate first order approximation.

$$e^{-\theta s} = \frac{1}{1 + \theta s} \tag{15.71}$$

The Padé approximation increases accuracy further, usually achieving that required using only first order terms. It comprises two series – one the numerator of a fraction, the other its denominator. The numerator is expressed as

$$P(x) = p_0 + p_1 x + p_2 x^2 + \ldots + p_m x^m . \tag{15.72}$$

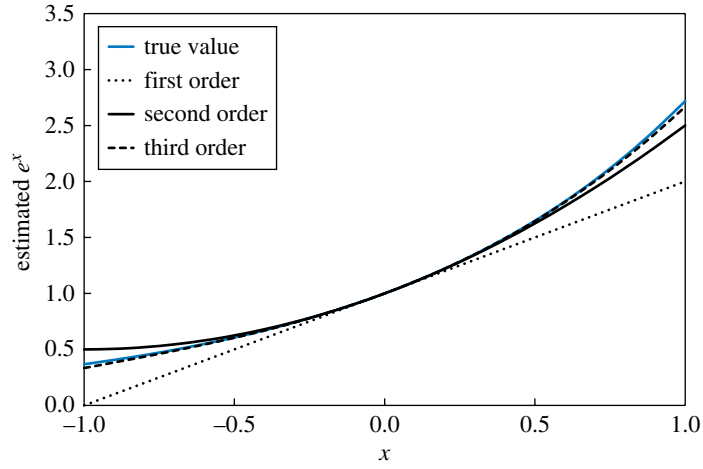

Figure 15.19 *Taylor approximation*

The denominator is

$$Q(x) = 1 + q_1 x + q_2 x^2 + \ldots + q_n x^n. \tag{15.73}$$

The method is to determine the coefficients p and q such that

$$f(x) = \frac{P(x)}{Q(x)} \quad \text{or} \quad f(x)Q(x) - P(x) = 0 \tag{15.74}$$

In our example $f(x)$ is the truncated Taylor series for e^x.

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^{m+n}}{(m+n)!} \tag{15.75}$$

We choose values for m and n to obtain the accuracy required. It is not a condition that we choose the same value for both but to explore the accuracy of a first order function we will choose a value of 1 for each. So, by combining Equations (15.72) to (15.75) and ignoring terms of order greater than $m+n$, we get

$$\left(1 + x + \frac{x^2}{2} \ldots\right)(1 + q_1 x) - (p_0 + p_1 x) = 0 \tag{15.76}$$

Rearranging

$$(1 - p_0) + (1 + q_1 - p_1)x + \left(\frac{1}{2} + q_1\right)x^2 = 0 \tag{15.77}$$

Since, in general, x will not be zero, we solve this equation as

$$p_0 = 1 \quad p_1 = \frac{1}{2} \quad q_1 = -\frac{1}{2} \tag{15.78}$$

And so

$$f(x)=\frac{P(x)}{Q(x)}=\frac{2+x}{2-x} \tag{15.79}$$

Some texts use a less general method of obtaining this equation by applying a first order Maclaurin approximation to both numerator and denominator as follows.

$$e^{x}=\frac{e^{x/2}}{e^{-x/2}}=\frac{1+\frac{x}{2}}{1-\frac{x}{2}}=\frac{2+x}{2-x} \tag{15.80}$$

As Figure 15.20 shows, this is almost as accurate as the third order Taylor approximation. Indeed, it is this function that is commonly used to approximate the Laplace transform for deadtime.

$$e^{-\theta s}=\frac{2-\theta s}{2+\theta s} \tag{15.81}$$

However, should we require greater accuracy, then with m and n both set at 2, we get

$$\left(1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}\ \ldots\ \right)\left(1+q_1x+q_2x^2\right)-\left(p_0+p_1x+p_2x^2\right)=0 \tag{15.82}$$

And so

$$\left(1-p_0\right)+\left(1+q_1-p_1\right)x+\left(\frac{1}{2}+q_1+q_2-p_2\right)x^2+\left(\frac{1}{6}+\frac{q_1}{2}+q_2\right)x^3+\left(\frac{1}{24}+\frac{q_1}{6}+\frac{q_2}{2}\right)x^4=0 \tag{15.83}$$

Solving gives

$$p_0=1\quad p_1=\frac{1}{2}\quad p_2=\frac{1}{12}\quad q_1=-\frac{1}{2}\quad q_2=\frac{1}{12} \tag{15.84}$$

And so

$$f(x)=\frac{P(x)}{Q(x)}=\frac{12+6x+x^2}{12-6x+x^2} \tag{15.85}$$

Thus

$$e^{-\theta s}=\frac{12-6\theta s+(\theta s)^2}{12+6\theta s+(\theta s)^2} \tag{15.86}$$

This function is also included in Figure 15.20, where it matches the real function almost perfectly. By the same method, if m is set to 0 and n to 1, we obtain Equation (15.71). If m is set to 1 and n to 0,

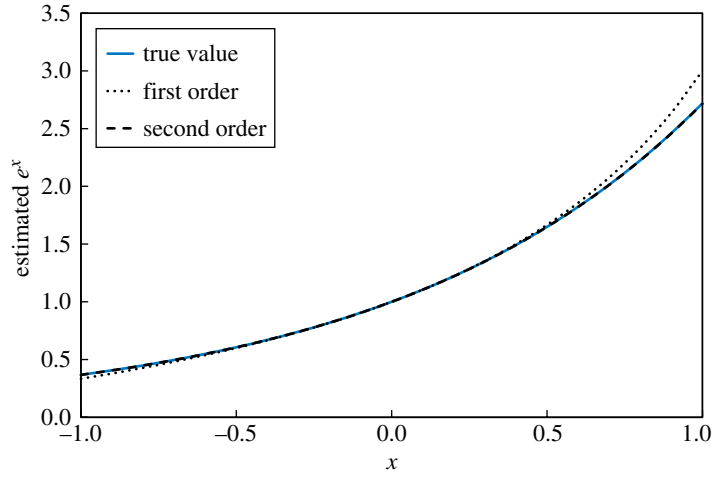

Figure 15.20 *Padé approximation*

we obtain Equation (15.69). In general, provided m is the same as n, the coefficients can be determined from

$$p_i = (-1)^i \frac{(2m-i)!m!}{(2m)!i!(m-i)!} \qquad q_i = \frac{(2m-i)!m!}{(2m)!i!(m-i)!} \tag{15.87}$$

If m and n are chosen not to be equal, then there is no simple formula and the calculation of coefficients must be made as above. Table 15.2 gives the coefficients for all combinations of m and n up to an order of 3. In all cases p_0 is 1. All the approximations can be refined by using least squares regression to optimise the coefficients. For example, doing so for the first order form, Equation (15.81), gives a function which is more accurate than all of those derived above.

$$e^{-\theta s} = \frac{2.028 - 1.048\theta s}{2 + 0.880\theta s} \tag{15.88}$$

15.5 Laplace Transform Derivations

If any one branch of mathematics has most daunted control engineers, it is the *Laplace transform*. For this reason we have largely avoided its use. But it can be a very convenient way of describing how processes and controllers behave. The control engineer should treat the transforms simply as a shorthand technique, without being too concerned about their theoretical basis. Here we present more of the background for those who wish to understand the technique in greater detail. The Laplace transform changes a function $f(t)$ written in the *time domain* to its equivalent $f(s)$ written in the *s-domain*. By definition, it is

$$f(s) = \int_0^\infty f(t).e^{-st}.dt \tag{15.89}$$

Table 15.2 *Padé coefficients for approximation of* $e^{-\theta^s}$

m	n	p_1	p_2	p_2	q_1	q_2	q_3
0	0						
	1				-1		
	2				-1	$\frac{1}{2}$	
	3				-1	$\frac{1}{2}$	$-\frac{1}{6}$
1	0	1					
	1	$\frac{1}{2}$			$-\frac{1}{2}$		
	2	$\frac{1}{3}$			$-\frac{2}{3}$	$\frac{1}{6}$	
	3	$\frac{1}{4}$			$-\frac{3}{4}$	$\frac{1}{4}$	$-\frac{1}{24}$
2	0	1	$\frac{1}{2}$				
	1	$\frac{2}{3}$	$\frac{1}{6}$		$-\frac{1}{3}$		
	2	$\frac{1}{2}$	$\frac{1}{12}$		$-\frac{1}{2}$	$\frac{1}{12}$	
	3	$\frac{2}{5}$	$\frac{1}{20}$		$-\frac{3}{5}$	$\frac{3}{20}$	$-\frac{1}{60}$
3	0	1	$\frac{1}{2}$	$\frac{1}{6}$			
	1	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{24}$	$-\frac{1}{4}$		
	2	$\frac{3}{5}$	$\frac{3}{20}$	$\frac{1}{60}$	$-\frac{2}{5}$	$\frac{1}{20}$	
	3	$\frac{1}{2}$	$\frac{1}{10}$	$\frac{1}{120}$	$-\frac{1}{2}$	$\frac{1}{10}$	$-\frac{1}{120}$

For example, a step change of magnitude 1 (*unit step*) made at $t = 0$ is described by $f(t) = 1$ and so

$$f(s) = \int_0^\infty e^{-st}.dt = \left[\frac{-e^{-st}}{s}\right]_0^\infty = \frac{1}{s} \tag{15.90}$$

Should the step be delayed by θ, then we integrate from this value rather than zero.

$$f(s) = \int_\theta^\infty e^{-st}.dt = \left[\frac{-e^{-st}}{s}\right]_\theta^\infty = \frac{e^{-\theta s}}{s} \tag{15.91}$$

The addition of the time delay results in the transform being multiplied by $e^{-\theta s}$ – the Laplace transform for a pure delay.

To understand how Laplace represents a lag, consider the exponential decay that is described by $f(t) = e^{-t/\tau}$.

$$f(s) = \int_0^\infty e^{-t/\tau} e^{-st}.dt = \left[\frac{-e^{-(s+1/\tau)t}}{s+1/\tau}\right]_0^\infty = \frac{1}{s+1/\tau} = \frac{\tau}{\tau s+1} \tag{15.92}$$

Many other functions can be derived, with perhaps only a few relevant to the control engineer. Some control schemes involve ramping. A ramp of slope 1, starting at $t = 0$ is described by $f(t) = t$, and so

$$f(s) = \int_0^\infty t.e^{-st}.dt \tag{15.93}$$

Remembering the general formula for integration by parts

$$\int u.dv = u.v - \int v.du \tag{15.94}$$

Where, in this case

$$u = t \qquad \text{and so} \qquad du = dt \tag{15.95}$$

$$dv = e^{-st}dt \qquad \text{and so} \qquad v = \frac{-e^{-st}}{s} \tag{15.96}$$

Substituting into Equation (15.94)

$$f(s) = \left[t\frac{-e^{-st}}{s}\right]_0^\infty - \int_0^\infty \frac{-e^{-st}}{s}dt = 0 - \left[\frac{e^{-st}}{s^2}\right]_0^\infty = \frac{1}{s^2} \tag{15.97}$$

Laplace transforms also have number of useful properties, for example

$$f_1(t) + f_2(t) = f_1(s) + f_2(s) \tag{15.98}$$

$$kf(t) = kf(s) \tag{15.99}$$

$$\frac{d\left(f\left(t\right)\right)}{dt} = sf\left(s\right) - f\left(0\right) \tag{15.100}$$

$$\int_0^t f\left(t\right).dt = \frac{f\left(s\right)}{s} \tag{15.101}$$

Using all of these properties we can convert the time-domain version of the conventional version of the PID algorithm to its *s*-domain equivalent.

$$M = K_c\left[E + \frac{1}{T_i}\int_0^t E.dt + T_d\frac{dE}{dt}\right] = K_c\left[1 + \frac{1}{T_i s} + T_d s\right]E \tag{15.102}$$

Similarly, combining Equations (2.6) and (2.7) gives the differential equation for a first order process with lag τ and zero deadtime.

$$\tau\frac{dx}{dt} + x = x_{new} \tag{15.103}$$

Modifying it to be based on *PV* and *MV* gives

$$\tau\frac{dPV}{dt} + PV = K_p MV \tag{15.104}$$

Applying Equation (15.100) gives the usual Laplace transform for a first order process with no deadtime.

$$\tau s.PV + PV = K_p MV \qquad \text{or} \qquad PV = \frac{K_p}{\tau s + 1}MV \tag{15.105}$$

To include deadtime (θ), as we saw with Equation (15.91), we multiply by $e^{-\theta s}$ to give the FOPDT transform

$$PV = \frac{K_p e^{-\theta s}}{\tau s + 1}MV \tag{15.106}$$

Most books covering the theory of process control will include a look-up table showing those transforms which are more difficult to derive. Table 15.3 is restricted to those in common use in process control. Also included are *z-transforms f(z)* that we cover in a later section.

15.6 Laplace Transforms for Processes

While this book only uses Laplace transforms when the alternative would be overly complex, they are used in many text books. So that they can be recognised, the transforms for the common types of process are listed here.

a. Self-regulating first order plus deadtime (FOPDT)

$$PV = \frac{K_p e^{-\theta s}}{\tau s + 1}MV \tag{15.107}$$

b. Integrating with deadtime

$$PV = \frac{K_p e^{-\theta s}}{s} MV \tag{15.108}$$

c. Self-regulating second order plus deadtime (SOPDT)

$$PV = \frac{K_p e^{-\theta s}}{(\tau_1 s + 1)(\tau_2 s + 1)} MV \tag{15.109}$$

A second order process can also be described by

$$PV = \frac{K_p e^{-\theta s}}{\tau^2 s^2 + 2\zeta\tau s + 1} MV \tag{15.110}$$

Table 15.3 *Common Laplace and z-transforms*

Description	$f(t)$	$f(s)$	$f(z)$
Unit step	1	$\frac{1}{s}$	$\frac{1}{1-z^{-1}}$
Unit ramp	t	$\frac{1}{s^2}$	$\frac{ts.z^{-1}}{\left(1-z^{-1}\right)^2}$
Delay	$f(t-\theta)=1$	$e^{-\theta s}$	
First order lag	$\frac{e^{-t/\tau}}{\tau}$	$\frac{1}{\tau s+1}$	$\frac{1}{\left(1-e^{-ts/\tau}z^{-1}\right)\tau}$
Integrator with lag	$1-e^{-t/\tau}$	$\frac{1}{s(\tau s+1)}$	$\frac{\left(1-e^{-ts/\tau}\right)z^{-1}\tau}{\left(1-z^{-1}\right)\left(1-e^{-ts/\tau}z^{-1}\right)}$
Second order ($\tau = \tau_1 = \tau_2$)		$\frac{1}{(\tau s+1)^2}$	$\frac{ts.e^{-ts/\tau}z^{-1}}{\left(1-e^{-ts/\tau}z^{-1}\right)^2\tau^2}$
Second order ($\tau_1 \neq \tau_2$)	$\frac{1}{\tau_1-\tau_2}\left(e^{-t/\tau_1}-e^{-t/\tau_2}\right)$	$\frac{1}{(\tau_1 s+1)(\tau_2 s+1)}$	$\frac{\tau_1\tau_2}{\tau_1-\tau_2}\left[\frac{1}{1-e^{-ts/\tau_1}z^{-1}}-\frac{1}{1-e^{-ts/\tau_2}z^{-1}}\right]$
Second order with lead	$\frac{1}{\tau_1}\frac{\tau_1-\tau_3}{\tau_1-\tau_2}e^{-t/\tau_1}+\frac{1}{\tau_2}\frac{\tau_2-\tau_3}{\tau_2-\tau_1}e^{-t/\tau_2}$	$\frac{\tau_3 s+1}{(\tau_1 s+1)(\tau_2 s+1)}$	
Zero order hold		$\frac{1-e^{-s.ts}}{s}$	

In Chapter 14 we introduced τ as the natural period of oscillation and ζ (zeta) as the dimensionless damping ratio. A second order process of this form can theoretically be separated into two first order processes. From Equations (15.109) and (15.110) we get

$$\tau^2 s^2 + 2\zeta\tau s + 1 = (1+\tau_1 s)(1+\tau_2 s) \tag{15.111}$$

Equating coefficients we get

$$\tau^2 = \tau_1\tau_2 \qquad \text{and} \qquad 2\zeta\tau = \tau_1 + \tau_2 \tag{15.112}$$

From which we obtain

$$\tau = \sqrt{\tau_1\tau_2} \qquad \text{and} \qquad \zeta = \frac{\tau_1 + \tau_2}{2\sqrt{\tau_1\tau_2}} \tag{15.113}$$

Alternatively we can factorise the left hand side of Equation (15.111)

$$\left(\frac{\tau s}{\zeta - \sqrt{\zeta^2 - 1}} + 1\right)\left(\frac{\tau s}{\zeta + \sqrt{\zeta^2 - 1}} + 1\right) = (1+\tau_1 s)(1+\tau_2 s) \tag{15.114}$$

So, providing ζ is not less than 1, arbitrarily choosing $\tau_1 > \tau_2$, then

$$\tau_1 = \frac{\tau}{\zeta - \sqrt{\zeta^2 - 1}} = \tau\left(\zeta + \sqrt{\zeta^2 - 1}\right) \tag{15.115}$$

$$\tau_2 = \frac{\tau}{\zeta + \sqrt{\zeta^2 - 1}} = \tau\left(\zeta - \sqrt{\zeta^2 - 1}\right) \tag{15.116}$$

If ζ is greater 1, then the process is overdamped. From the calculations performed by others [47], pp 115–116, the response to a unit step change in the MV can then be described by

$$PV = K_p\left[1 - \frac{\tau_1 e^{-t/\tau_1} - \tau_2 e^{-t/\tau_2}}{\tau_1 - \tau_2}\right]MV \tag{15.117}$$

Or, if left unfactorised

$$PV = K_p\left[1 - e^{-\zeta t/\tau}\left[\cosh\left(\frac{\sqrt{\zeta^2 - 1}}{\tau}t\right) + \frac{\zeta}{\sqrt{\zeta^2 - 1}}\sinh\left(\frac{\sqrt{\zeta^2 - 1}}{\tau}t\right)\right]\right]MV \tag{15.118}$$

If ζ is equal to 1, then $\tau_1 = \tau_2$ (= τ) and the process is critically damped. The response can then be described by

$$PV = K_p\left[1 - \left(1 + \frac{t}{\tau}\right)e^{-t/\tau}\right]MV \tag{15.119}$$

For values of ζ greater than 0 and less than 1 the process is underdamped. In the process industry there are very few underdamped processes, but underdamped behaviour can be the result of adding a poorly tuned controller. What we have so far called lags, τ_1 and τ_2, cannot now be expressed as real numbers. The response can be described as

$$PV = K_p\left[1 - e^{-\zeta t/\tau}\left[\cos\left(\frac{\sqrt{1-\zeta^2}}{\tau}t\right) + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\left(\frac{\sqrt{1-\zeta^2}}{\tau}t\right)\right]\right]MV \tag{15.120}$$

The process response (in this case with ζ set to 0.15) is shown in Figure 15.21. In order to determine the first and second PV overshoots (A and B) we differentiate Equation (15.120) and set dPV/dt to zero.

$$t = \frac{n\pi\tau}{\sqrt{1-\zeta^2}} \tag{15.121}$$

The integer term n is even for troughs in the oscillatory response and odd for peaks. The first peak occurs therefore when n is 1. Substituting Equation (15.121) into Equation (15.120) gives

$$PV = K_p\left[1 + \exp\left(\frac{-\pi\zeta}{\sqrt{1-\zeta^2}}\right)\right]MV \tag{15.122}$$

Thus, the PV overshoot, expressed as a fraction of the steady state change, is given by

$$A = \exp\left(\frac{-\pi\zeta}{\sqrt{1-\zeta^2}}\right) \tag{15.123}$$

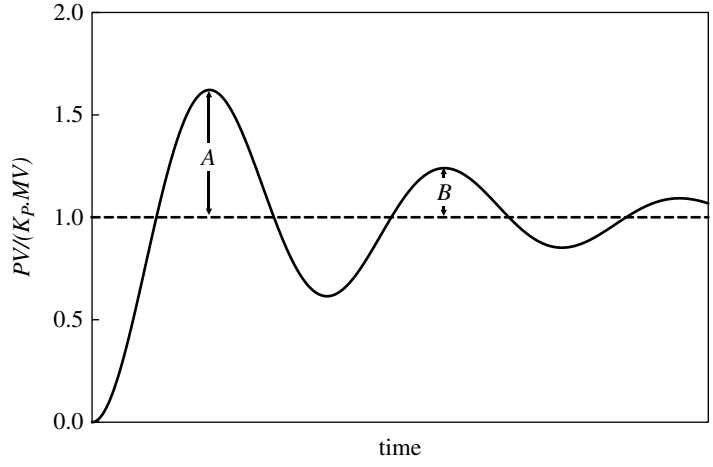

Figure 15.21 *Underdamped response*

The second peak occurs when n is 3; here the PV overshoot is given by

$$B = \exp\left(\frac{-3\pi\zeta}{\sqrt{1-\zeta^2}}\right) \tag{15.124}$$

The decay ratio can be determined by dividing Equation (15.124) by Equation (15.123).

$$\frac{B}{A} = \exp\left(\frac{-2\pi\zeta}{\sqrt{1-\zeta^2}}\right) \tag{15.125}$$

The effect that ζ has on PV overshoot and decay ratio is illustrated in Figure 15.22. For example to achieve the quarter decay ratio of 0.25, suggested by the Ziegler-Nichols and Cohen-Coon tuning methods, ζ should be 0.215 – remembering that the second order model will now include both process and controller.

If ζ has a value of 0, the process is undamped and the oscillation will not decay and so will persist with constant amplitude. The response can be derived from Equation (15.120) and is described by

$$PV = K_p\left[1-\cos\left(\frac{t}{\tau}\right)\right]MV \tag{15.126}$$

From this equation we can see that the frequency (f) of oscillation is given by $1/\tau$ (radians per unit time). The period of oscillation is given by $1/f$ or τ – hence it being described as the natural period (in seconds per radian). More usually the period is expressed as the time for a complete cycle (2π radians) and so is given by $2\pi\tau$.

If ζ has a value less than 0, the process is unstable and the amplitude oscillation will increase over time. The response is described by Equation (15.120).

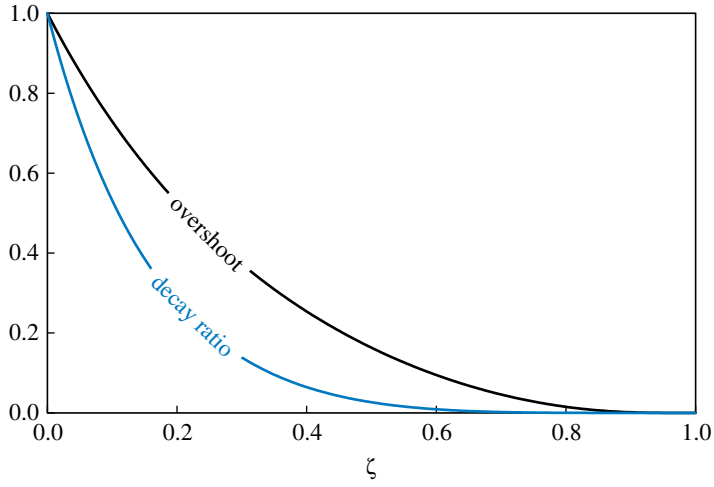

Figure 15.22 *Effect of ζ on PV overshoot and decay ratio*

The response curves for a range of damping ratios are plotted, using Equations (15.117), (15.119) and (15.120), in Figure 15.23.

d. Self-regulating second order with inverse response

As shown in Chapter 2, inverse response is caused by two competing processes – the faster of which takes the process first in a direction opposite to the steady state. We can approximate this as two first-order processes with gains of opposite sign, so that the combined effect is given by

$$PV = \left[\frac{(K_p)_1}{\tau_1 s + 1} + \frac{(K_p)_2}{\tau_2 s + 1}\right] MV \tag{15.127}$$

Rearranging

$$PV = \frac{\left((K_p)_1 + (K_p)_2\right)\left[\dfrac{(K_p)_1 \tau_2 + (K_p)_2 \tau_1}{(K_p)_1 + (K_p)_2} s + 1\right]}{(\tau_1 s + 1)(\tau_2 s + 1)} MV \tag{15.128}$$

Defining

$$K_p = (K_p)_1 + (K_p)_2 \tag{15.129}$$

and

$$\tau_3 = \frac{(K_p)_1 \tau_2 + (K_p)_2 \tau_1}{(K_p)_1 + (K_p)_2} \tag{15.130}$$

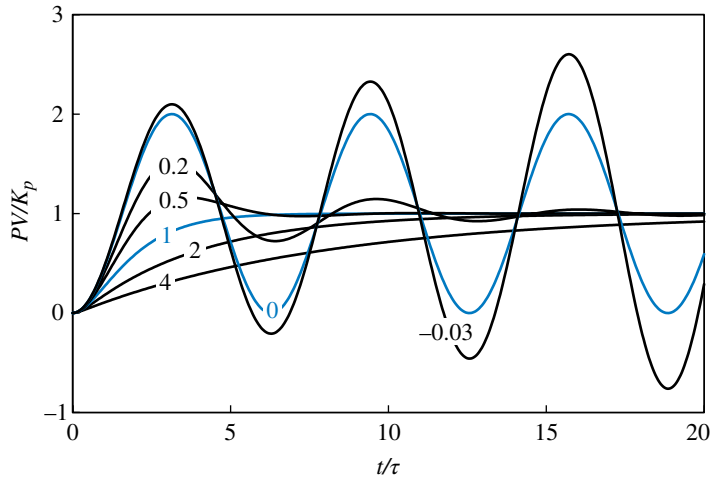

Figure 15.23 *Effect of damping ratio on process response*

We can write Equation (15.128) as

$$PV = \frac{K_p\left(\tau_3 s+1\right)}{\left(\tau_1 s+1\right)\left(\tau_2 s+1\right)} MV \tag{15.131}$$

If τ_3 is less than zero, then the process will show inverse response. If τ_3 is greater than zero, then the process will show PV overshoot and τ_3 is said to add lead to the process.

While useful in understanding the Laplace form of a second order process, we have assumed that neither of the original first order processes have deadtime. Deadtime can of course be included in Equation (15.131). Indeed, most model identification packages that have the capability to fit a second model will identify one having the form given by Equation (15.132).

$$PV = \frac{K_p e^{-\theta s}\left(\tau_3 s+1\right)}{\left(\tau_1 s+1\right)\left(\tau_2 s+1\right)} MV \tag{15.132}$$

Strictly this is only correct if the deadtimes of each of the first order processes are equal (in this case to θ). Since this is not usually the case, then a more accurate representation is given by Equation (15.133).

$$PV = \left[\frac{\left(K_p\right)_1 e^{-\theta_1 s}}{\tau_1 s+1} + \frac{\left(K_p\right)_2 e^{-\theta_2 s}}{\tau_2 s+1}\right] MV \tag{15.133}$$

15.7 Laplace Transforms for Controllers

Except where the alternative discrete formula was too complex, we have been able to avoid using Laplace transforms to describe controllers. But there are many text books that use them throughout. Further, many of the DCS vendors document their system in this way – which strictly is incorrect since they should only be used to describe analog control. Further, in most DCS, controllers are implemented in the velocity form rather than the full position form described by Laplace. However, so that the engineers can recognise them, the transforms for the common types of controller are listed here – along with their time domain equivalents.

a. Ideal PID

$$M = K_c\left(E + \frac{1}{T_i}\int E.dt + T_d\frac{dE}{dt}\right) \tag{15.134}$$

$$M = K_c\left[1 + \frac{1}{T_i s} + T_d s\right]E \tag{15.135}$$

b. Ideal PID with derivative gain limit

$$M = K_c\left[1 + \frac{1}{T_i s} + \frac{T_d s}{1+aT_d s}\right]E \tag{15.136}$$

c. Ideal PID with derivative-on-PV (PI-D)

$$M = K_c\left(E + \frac{1}{T_i}\int E.dt + T_d\frac{dPV}{dt}\right) \tag{15.137}$$

$$M = K_c\left(\left[1 + \frac{1}{T_i s}\right]E + T_d s PV\right) \tag{15.138}$$

$$M = K_c\left(\left[1 + \frac{1}{T_i s} + T_d s\right]PV - \left[1 + \frac{1}{T_i s}\right]SP\right) \tag{15.139}$$

d. Ideal PID with proportional-on-PV and derivative-on-PV (I-PD)

$$M = K_c\left(PV + \frac{1}{T_i}\int E.dt + T_d\frac{dPV}{dt}\right) \tag{15.140}$$

$$M = K_c\left(\frac{1}{T_i s}E + \left[1 + T_d s\right]PV\right) \tag{15.141}$$

$$M = K_c\left(\left[1 + \frac{1}{T_i s} + T_d s\right]PV - \left[\frac{1}{T_i s}\right]SP\right) \tag{15.142}$$

e. Interactive PID

$$M = K_c\left(\left(1 + \frac{T_d}{T_i}\right)E + \frac{1}{T_i}\int E.dt + T_d\frac{dE}{dt}\right) \tag{15.143}$$

$$M = K_c\left[1 + \frac{1}{T_i s}\right]\left[1 + T_d s\right]E \tag{15.144}$$

f. Interactive PID (with derivative-on-PV)

$$M = K_c\left(E + \frac{T_d}{T_i}PV + \frac{1}{T_i}\int E.dt + T_d\frac{dPV}{dt}\right) \tag{15.145}$$

$$M = K_c\left(\left[1 + \frac{1}{T_i s}\right]E + \left[\frac{T_d}{T_s} + T_d s\right]PV\right) \tag{15.146}$$

$$M = K_c\left(\left[1 + \frac{1}{T_i s}\right]\left[1 + T_d s\right]PV - \left[1 + \frac{1}{T_i s}\right]SP\right) \tag{15.147}$$

g. Interactive PID (with proportional-on-PV and derivative-on-PV)

$$M = K_c\left(\left(1 + \frac{T_d}{T_i}\right)PV + \frac{1}{T_i}\int E.dt + T_d\frac{dPV}{dt}\right) \tag{15.148}$$

$$M = K_c\left(\frac{1}{T_i s}E + \left[1 + \frac{T_d}{T_i} + T_d s\right]PV\right) \tag{15.149}$$

$$M = K_c\left(\left[1 + \frac{1}{T_i s}\right]\left[1 + T_d s\right]PV - \frac{1}{T_i s}SP\right) \tag{15.150}$$

h. Interactive PID with derivative gain limit

$$M = K_c\left[1 + \frac{1}{T_i s}\right]\left[\frac{1 + T_d s}{1 + aT_d s}\right]E \tag{15.151}$$

i. Interactive PID with derivative gain limit (with derivative-on-PV)

$$M = K_c\left(\left[1 + \frac{1}{T_i s}\right]\left[\frac{1 + T_d s}{1 + aT_d s}\right]PV - \left[1 + \frac{1}{T_i s}\right]SP\right) \tag{15.152}$$

j. Interactive PID with derivative gain limit (with proportional-on-PV and derivative-on-PV)

$$M = K_c\left(\left[1 + \frac{1}{T_i s}\right]\left[\frac{1 + T_d s}{1 + aT_d s}\right]PV - \frac{1}{T_i s}SP\right) \tag{15.153}$$

k. Integral only

$$M = \frac{1}{T_i s}E \tag{15.154}$$

An alternative method of implementing the integral part of the PID algorithm may be used when there is a need for reset feedback (sometimes known as external reset feedback). Instead of integral action being based on the input to the controller, it is based on applying a lag (with time constant T_i) to the output. Applying this, for example to the interactive algorithm, gives the structure of the controller as shown in Figure 15.24. Its transfer function is given by Equation (15.155) – which rearranges to give Equation (15.144).

$$M = K_c\left(1 + T_d s\right)E + \left(\frac{1}{T_i s + 1}\right)M \tag{15.155}$$

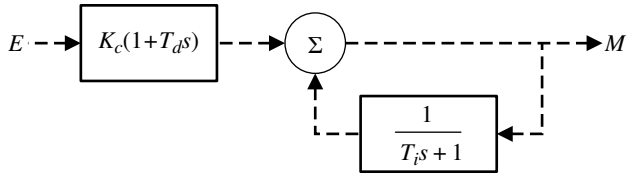

Figure 15.24 *Alternative interactive PID algorithm structure*

Basing integral action on the actual output rather than error provides anti-reset windup in full position controllers. This is particularly useful if the controller is one of two or more outputting to a signal selector. Basing all the controllers on the selected output prevents the deselected controller(s) from winding up.

15.8 I-PD versus PI-D Algorithm

We described, in Chapter 3, the benefit of applying the proportional-on-PV (I-PD) version of the control algorithm compared to the more conventional proportional-on-error (PI-D) version. For the same tuning they give very different responses to SP changes. However, by applying suitable conditioning to the SP, the PI-D version can be made to perform identically to the I-PD version. This technique would be used if the I-PD algorithm is not supported by the control system.

If the conditioned SP is SP^* then we can equate Equations (15.138) and (15.141) to give, with identical tuning, the same response to a SP change.

$$K_c\left(\left[1+\frac{1}{T_i s}\right]\left(PV-SP^*\right)+T_d s PV\right)=K_c\left(\frac{1}{T_i s}\left(PV-SP\right)+\left[1+T_d s\right]PV\right) \tag{15.156}$$

Solving gives

$$SP^*=\frac{SP}{T_i s+1} \tag{15.157}$$

This is a conventional first order exponential filter with a filter lag of T_i. So, by applying this filter to the SP, we can tune the algorithm as if it is the I-PD type and so benefit from the improved response to load changes, without the problem of the controller acting too aggressively when the SP is changed. We can take the same approach for digital control by modifying Equation (3.39) (to make it proportional-on-PV) and equating it with Equation (3.82).

$$\begin{aligned}&K_c\left[\left(PV_n-SP_n^*\right)-\left(PV_{n-1}-SP_{n-1}^*\right)+\frac{ts}{T_i}\left(PV_{n-1}-SP_{n-1}^*\right)+\frac{T_d}{ts}\left(PV_n-2PV_{n-1}+PV_{n-2}\right)\right]\\&=K_c\left[\left(PV_n-PV_{n-1}\right)+\frac{ts}{T_i}\left(PV_{n-1}-SP_{n-1}\right)+\frac{T_d}{ts}\left(PV_n-2PV_{n-1}+PV_{n-2}\right)\right]\end{aligned} \tag{15.158}$$

Solving gives

$$SP_n^*=\left(1-\frac{ts}{T_i}\right)SP_{n-1}^*+\left(\frac{ts}{T_i}\right)SP_{n-1} \tag{15.159}$$

Comparison with Equations (5.29) and (5.31) shows that this is also an exponential filter, with lag set to T_i, but one based on the previous SP. However, using the more common version of the PID algorithm, described by Equation (3.39), gives a slightly different result – conventionally based on the current SP.

$$SP_n^*=\left(\frac{T_i}{T_i+ts}\right)SP_{n-1}^*+\left(\frac{ts}{T_i+ts}\right)SP_n \tag{15.160}$$

However, since ts is usually much smaller than T_i, a conventional filter will be a very close approximation.

Normally β, in the two degrees of freedom controller described by Equations (3.83) and (3.84), is set to zero. Converting to Laplace and equating to Equation (15.138) gives

$$K_c\left(\left[1+\frac{1}{T_i s}\right]\left(PV-SP^*\right)+T_d sPV\right)=K_c\left(\left(PV-\alpha SP\right)+\frac{1}{T_i s}\left(PV-SP\right)+T_d sPV\right) \tag{15.161}$$

Solving gives

$$SP^* = \frac{\alpha T_i s+1}{T_i s+1} SP \tag{15.162}$$

The conditioning is now a lead-lag with lead set to αT_i and lag set to T_i. The term α is thus the lead-to-lag ratio. As illustrated by Equation (15.157), setting it to zero causes the PI-D algorithm to replicate the response of the I-PD algorithm. Setting it to 1 has no effect on the response, while setting it between these two values replicates the behaviour of the two degrees of freedom controller.

15.9 Direct Synthesis

Direct synthesis is a theoretical method of deriving the optimum tuning for a controller. It is the technique used to derive the formulae for both the Lambda and IMC tuning methods described in Chapter 3. Neither of these methods are recommended; indeed, nor are any that are derived using direct synthesis. Its inclusion here is primarily to demonstrate the limitations of the method and so show, in detail, the reasoning behind the recommendation. If the reader has already accepted these conclusions, then the method is of little more than academic interest. But readers, not daunted by the mathematics, should find it of value in understanding the derivation of the wide range of IMC tuning formulae published elsewhere.

Figure 15.25 shows the control loop, where G_c is the transfer function of the controller and G_p is the transfer function for the process. PV is related to E by the combination of the controller and process transfer functions

$$PV = G_c G_p E \tag{15.163}$$

But, by definition

$$E = PV - SP \tag{15.164}$$

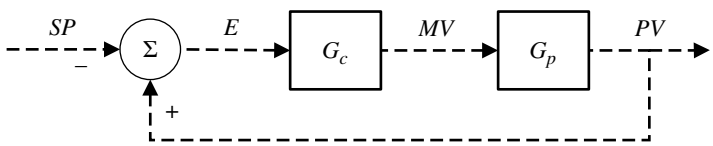

Figure 15.25 Control loop

Substituting this in Equation (15.163) gives

$$PV = \frac{G_c G_p}{G_c G_p - 1} SP \tag{15.165}$$

This describes how the PV responds to changes in SP. We wish this trajectory to be a first order response with a lag of λ, where the value of λ is selected by the engineer. Therefore

$$\frac{G_c G_p}{G_c G_p - 1} = \frac{1}{1 + \lambda s} \qquad \text{or} \qquad G_c = \frac{-1}{G_p \lambda s} \tag{15.166}$$

Let us assume that we have a simple first order process with no deadtime, then

$$G_p = \frac{K_p}{1 + \tau s} \tag{15.167}$$

Substituting into Equation (15.166)

$$G_c = \frac{-(1 + \tau s)}{K_p \lambda s} = \frac{-\tau}{K_p \lambda}\left[1 + \frac{1}{\tau s}\right] \tag{15.168}$$

Comparing this to Equation (15.135), this is a PID controller with

$$K_c = \frac{-\tau}{K_p \lambda} \qquad T_i = \tau \qquad T_d = 0 \tag{15.169}$$

K_c has an opposite sign to K_p because, according to our definition of error, a reverse-acting controller is required if the process gain is positive. The result of this method produced a transfer function identical to that of a PID algorithm. Rarely does this occur with other process models, for example, especially if we introduce deadtime into the process. The controller cannot cause the PV to change any sooner than the process deadtime (θ) permits. Our reference trajectory is therefore based on the same deadtime.

$$\frac{G_c G_p}{G_c G_p - 1} = \frac{e^{-\theta s}}{1 + \lambda s} \qquad \text{or} \qquad G_c = \frac{-e^{-\theta s}}{G_p\left(1 + \lambda s - e^{-\theta s}\right)} \tag{15.170}$$

And the process transfer function becomes

$$G_p = \frac{K_p e^{-\theta s}}{1 + \tau s} \tag{15.171}$$

Substituting into Equation (15.170)

$$G_c = \frac{-(1 + \tau s)}{K_p\left(1 + \lambda s - e^{-\theta s}\right)} \tag{15.172}$$

The presence of $e^{-\theta s}$ means that the result is not a PID controller. We can resolve this by making the first order Taylor approximation

$$e^{-\theta s} = 1 - \theta s \tag{15.173}$$

Substituting into Equation (15.172)

$$G_c = \frac{-(1+\tau s)}{K_p(\lambda+\theta)s} = \frac{-\tau}{K_p(\lambda+\theta)}\left[1+\frac{1}{\tau s}\right] \tag{15.174}$$

This is a PID controller, tuned according to the Lambda tuning method, as described in Section 3.14.

$$K_c = \frac{-\tau}{K_p(\lambda+\theta)} \qquad T_i = \tau \qquad T_d = 0 \tag{15.175}$$

As we saw earlier in this chapter, a more accurate first order approximation is possible.

$$e^{-\theta s} = \frac{1}{1+\theta s} \tag{15.176}$$

We also saw, in Chapter 2, that deadtime can be approximated by a series of n first order lags. Each lag then will be θ/n.

$$e^{-\theta s} = \frac{1}{\left(1+\frac{\theta}{n}s\right)^n} \tag{15.177}$$

The first order approximation ($n = 1$) of Equation (15.177) is the same the first order Taylor approximation given by Equation (15.176). However, using this approximation, results in a controller of a form which cannot be replicated by the PID algorithm. Being forced to use a less accurate approximation must place some suspicion on the reliability of the tuning determined by Equation (15.175). Indeed, this is confirmed by the fact the technique results in zero derivative action, no matter what the deadtime. A better approach is to make the first order Padé approximation

$$e^{-\theta s} = \frac{2-\theta s}{2+\theta s} \tag{15.178}$$

Substituting into Equation (15.172)

$$G_c = \frac{-(1+\tau s)(2+\theta s)}{K_p\left[(1+\lambda s)(2+\theta s)-(2-\theta s)\right]} = \frac{-\left[2+(2\tau+\theta)s+\theta\tau s^2\right]}{K_p\left[(2\lambda+2\theta)s+\theta\lambda s^2\right]} \tag{15.179}$$

Since they cannot be part of a PID algorithm, we ignore high order terms of s^2 and above.

$$G_c = \frac{-2-(2\tau+\theta)s}{K_p(2\lambda+2\theta)s} = \frac{-\left(\tau+\theta/2\right)}{K_p(\lambda+\theta)}\left[1+\frac{1}{\left(\tau+\theta/2\right)s}\right] \tag{15.180}$$

This is a PID controller with

$$K_c = \frac{-\left(\tau + \theta/2\right)}{K_p\left(\lambda + \theta\right)} \qquad T_i = \tau + \theta/2 \qquad T_d = 0 \tag{15.181}$$

This too results in zero derivative action, again placing suspicion on the result. An improvement is to ignore the high order terms only in the denominator of Equation (15.179).

$$G_c = \frac{-\left[2 + \left(2\tau + \theta\right)s + \theta\tau s^2\right]}{K_p\left(2\lambda + 2\theta\right)s} = \frac{-\left(\tau + \theta/2\right)}{K_p\left(\lambda + \theta\right)}\left[1 + \frac{1}{\left(\tau + \theta/2\right)s} + \frac{\theta\tau s}{2\tau + \theta}\right] \tag{15.182}$$

This is a PID controller, tuned as described in Table 3.7.

$$K_c = \frac{-\left(\tau + \theta/2\right)}{K_p\left(\lambda + \theta\right)} \qquad T_i = \tau + \theta/2 \qquad T_d = \frac{\theta\tau}{2\tau + \theta} \tag{15.183}$$

In summary, if we compare this result with Equations (15.175) and (15.181), we get slightly different tuning formulae – depending on how the approximations are made. Yet another approach to the same process is possible. Instead of ignoring terms in s^2 we can rearrange Equation (15.179).

$$G_c = \frac{-\left(\tau + \theta/2\right)}{K_p\left(\lambda + \theta\right)}\left[1 + \frac{1}{\left(\tau + \theta/2\right)s} + \frac{\theta\tau s}{2\tau + \theta}\right]\left[\frac{1}{1 + \dfrac{\lambda\tau s}{2\left(\lambda + \theta\right)}}\right] \tag{15.184}$$

This gives the same PID tuning as Equation (15.183) but now includes a filter with a lag time constant of (τ_f).

$$\tau_f = \frac{\lambda\tau}{2\left(\lambda + \theta\right)} \tag{15.185}$$

However, the filter is applied to the SP not the PV. Any filter placed on the PV will be taken account of in the value for the process dynamics (θ and τ) obtained from plant testing. One might argue that lagging SP changes is a decidedly odd thing to do, particularly if we wish to achieve a fast return to SP, simply for the trajectory to exactly follow that of a first order lag.

Another approach is to develop, from Equation (15.172), a controller of a form quite different from the PID algorithm.

$$M = \frac{-\left(1 + \tau s\right)}{K_p\left(1 + \lambda s - e^{-\theta s}\right)}E \quad \text{or} \quad M = -\frac{1}{K_p}\frac{1 + \tau s}{1 + \lambda s}E + \frac{e^{-\theta s}}{1 + \lambda s}M \tag{15.186}$$

This controller is shown in Figure 15.26. Of note is its similarity to the IMC technique described in Chapter 7. Indeed, it explains why the term 'IMC' can be used to describe both a PID tuning method and a controller in its own right. If λ is set equal to τ the lead-lag algorithm is effectively removed.

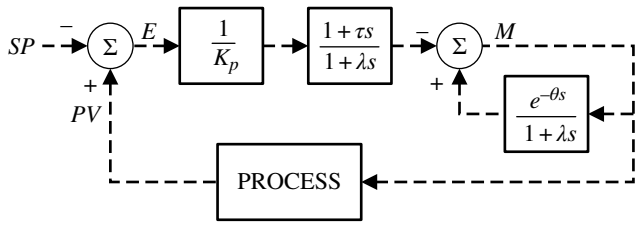

Figure 15.26 *Internal model controller*

The controller then becomes very similar to that shown in Figure 15.24 but with deadtime (θ) added to the integral action. If we set K_c to $1/K_p$, T_i to τ and T_d to zero, the configuration is then known as the *predictive PI controller*.

The direct synthesis technique may be applied to any process. A second order process without deadtime can be described in terms of damping ratio (ζ_p) by

$$G_p = \frac{K_p}{\tau^2 s^2 + 2\zeta_p \tau s + 1} \tag{15.187}$$

Our reference trajectory is a second order response with time constant λ and damping ratio ζ_c. From Equation (15.165)

$$\frac{G_c G_p}{G_c G_p - 1} = \frac{1}{\lambda^2 s^2 + 2\zeta_c \lambda s + 1} \quad \text{or} \quad G_c = \frac{-1}{G_p \lambda s\left(\lambda s + 2\zeta_c\right)} \tag{15.188}$$

Substituting from Equation (15.187)

$$G_c = \frac{-\left(\tau^2 s^2 + 2\zeta_p \tau s + 1\right)}{K_p \lambda s\left(\lambda s + 2\zeta_c\right)} \tag{15.189}$$

Ignoring terms in s^2 in the denominator, this can be rearranged.

$$G_c = \frac{-\zeta_p \tau}{K_p \zeta_c \lambda}\left[1 + \frac{1}{2\zeta_p \tau s} + \frac{\tau s}{2\zeta_p}\right] \tag{15.190}$$

This is a PID controller with

$$K_c = \frac{-\zeta_p \tau}{K_p \zeta_c \lambda} \qquad T_i = 2\zeta_p \tau \qquad T_d = \frac{\tau}{2\zeta_p} \tag{15.191}$$

As described in Chapter 3, a disadvantage of this tuning method is that it requires the engineer to select (usually by trial and error) a value for λ. If using a second order model, then a value for ζ_c must also be chosen – although this can reasonably be set to 1. Again, including a filter on the SP would mean we do not need to exclude terms in s^2; the filter lag (τ_f) would be given by

$$\tau_f = \frac{\lambda}{2\zeta_c} \tag{15.192}$$

A similar approach can be taken for integrating processes. For example, one with deadtime is described by

$$G_p = \frac{K_p e^{-\theta s}}{s} \tag{15.193}$$

Substituting into Equation (15.170)

$$G_c = \frac{-s}{K_p\left(1+\lambda s - e^{-\theta s}\right)} \tag{15.194}$$

Applying the first order Taylor approximation to $e^{-\theta s}$ gives

$$G_c = \frac{-1}{K_p\left(\lambda+\theta\right)} \tag{15.195}$$

This is proportional only controller with

$$K_c = \frac{-1}{K_p\left(\lambda+\theta\right)} \tag{15.196}$$

This nicely illustrates a limitation of direct synthesis. While such a controller will ensure the PV will follow a change in SP, there will be an offset if there is a load change.

Instead of applying the Taylor approximation, we use Padé and so Equation (15.195) becomes

$$G_c = \frac{-s\left(2+\theta s\right)}{K_p\left[\left(1+\lambda s\right)\left(2+\theta s\right)-\left(2-\theta s\right)\right]} = \frac{-\left(2+\theta s\right)}{K_p\left(2\lambda+2\theta+\lambda\theta s\right)} \tag{15.197}$$

This can be rewritten

$$G_c = \frac{-1}{K_p\left(\lambda+\theta\right)}\left[1+\frac{\theta}{2}s\right]\left[\frac{1}{1+\dfrac{\lambda\theta}{2\left(\lambda+\theta\right)}s}\right] \tag{15.198}$$

This is a PD controller with a filter on the SP.

$$K_c = \frac{-1}{K_p\left(\lambda+\theta\right)} \qquad T_d = \frac{\theta}{2} \qquad \tau_f = \frac{\lambda\theta}{2\left(\lambda+\theta\right)} \tag{15.199}$$

While derivative action has now been introduced, there is still no integral action and so an offset will result from load changes.

15.10 Predicting Filter Attenuation

A topic commonly included in process control courses taught at universities is the *frequency response* of processes (and controllers). This is determined by injecting a sine wave and determining how it is changed by the process in terms of both amplitude and phase. This then generally leads to the use of a *Bode plot*,

Nyquist diagram or *Nichols chart*. These techniques have little application in the process industry and perhaps should be excluded from chemical engineering courses. However, there are two areas where the frequency response can be of value. The first, described in this section, is to predict the behaviour of filters. To do so we have to assume that the noise signal is sinusoidal. Since this is rarely the case, it means that this provides only an approximate prediction of the filter performance. The second application, predicting controller stability limits, we will cover in the next section.

A first order exponential filter, with lag τ_f, can be represented in its Laplace form as

$$G_f(s) = \frac{1}{\tau_f s + 1} \tag{15.200}$$

Applying the shortcut method for finding the frequency response [47], p 317, we replace s with $j\omega$, where j is $\sqrt{-1}$ and ω is the noise frequency (in radians/minute). We thus convert a Laplace transform into its equivalent in the *frequency domain*.

$$G_f(j\omega) = \frac{1}{\tau_f j\omega + 1} \tag{15.201}$$

Multiplying both the numerator and denominator by $(\tau_f j\omega - 1)$ gives

$$G_f(j\omega) = \frac{\tau_f j\omega - 1}{\tau_f^2 \omega^2 + 1} \tag{15.202}$$

Splitting into its real (R) and imaginary parts (I)

$$G_f(j\omega) = R + j.I \tag{15.203}$$

where

$$R = \frac{-1}{\tau_f^2 \omega^2 + 1} \tag{15.204}$$

$$I = \frac{\tau_f \omega}{\tau_f^2 \omega^2 + 1} \tag{15.205}$$

The attenuation K_f (or amplitude ratio AR) is then given by

$$AR = \sqrt{R^2 + I^2} \tag{15.206}$$

$$K_f = \frac{1}{\sqrt{\tau_f^2 \omega^2 + 1}} \tag{15.207}$$

If τ_f has units of minutes, the noise frequency (f) will be measured in min^{-1}. Replacing ω with $2\pi f$ gives Equation (5.37).

15.11 Stability Limit for PID Control

It is possible to predict, from the process dynamics and PID tuning constants, at what point the control will become oscillatory as the process dynamics change. This can be used to check on the robustness of the controller tuning. By applying this technique to a proportional-only controller it can also be used to predict

the Ziegler-Nichols ultimate gain (K_u) and ultimate period of oscillation (P_u). Prediction relies on the fact that, at the limit of stability, the oscillation is sinusoidal with the controller output amplitude equal to the input amplitude. At this point the input and output are also in exact *anti-phase*, i.e. the phase shift is π radians (180°). The point at which this occurs is known as the *crossover frequency*.

Assuming a FOPDT self-regulating process

$$G_p(s) = \frac{K_p e^{-\theta s}}{\tau s + 1} \tag{15.208}$$

Applying the same method for finding the frequency response as used in the previous section

$$G_p(j\omega) = \frac{K_p e^{-\theta j\omega}}{\tau j\omega + 1} \tag{15.209}$$

Multiplying both the numerator and denominator by ($\tau j\omega - 1$) gives

$$G_p(j\omega) = \frac{K_p e^{-\theta j\omega}(\tau j\omega - 1)}{\tau^2\omega^2 + 1} \tag{15.210}$$

Euler's formula tells us that

$$e^{-\theta j\omega} = \cos(-\theta\omega) + j.\sin(-\theta\omega) \tag{15.211}$$

And so

$$G_p(j\omega) = \frac{K_p}{\tau^2\omega^2 + 1}\left[(\tau j\omega - 1).\cos(-\theta\omega) - (\tau\omega + j).\sin(-\theta\omega)\right] \tag{15.212}$$

The PID controller has the transfer function

$$G_c(s) = K_c\left[1 + \frac{1}{T_i s} + T_d s\right] \tag{15.213}$$

Making the same substitution

$$G_c(j\omega) = K_c\left[1 + \frac{1}{T_i j\omega} + T_d j\omega\right] \tag{15.214}$$

Multiplying the numerator and denominator of the integral term by j

$$G_c(j\omega) = K_c\left[1 - \frac{j}{T_i\omega} + T_d j\omega\right] \tag{15.215}$$

Multiplying G_p by G_c gives the overall transfer function which can then be split it into its real (R) and imaginary (I) parts

$$G_p(j\omega).G_c(j\omega) = R + j.I \tag{15.216}$$

where

$$R = \frac{K_p K_c}{\tau^2\omega^2 + 1}\left[A.\sin(-\theta\omega) + B.\cos(-\theta\omega)\right] \tag{15.217}$$

$$I = \frac{K_p K_c}{\tau^2 \omega^2 + 1}\left[B.\sin\left(-\theta\omega\right) - A.\cos\left(-\theta\omega\right)\right] \tag{15.218}$$

$$A = T_d\omega - \frac{1}{T_i\omega} - \tau\omega \tag{15.219}$$

$$B = \frac{\tau}{T_i} - \tau T_d\omega^2 - 1 \tag{15.220}$$

Applying Equation (15.206), replacing R and I as defined by Equations (15.217) and (15.218) and remembering that

$$\sin^2\left(-\theta\omega\right) + \cos^2\left(-\theta\omega\right) = 1 \tag{15.221}$$

gives

$$AR = \frac{K_p K_c}{\tau^2 \omega^2 + 1}\sqrt{A^2 + B^2} \tag{15.222}$$

The phase lag is (ϕ) given by

$$\tan\left(\phi\right) = \left(\frac{I}{R}\right) \tag{15.223}$$

Remembering that

$$\frac{\sin\left(-\theta\omega\right)}{\cos\left(-\theta\omega\right)} = \tan\left(-\theta\omega\right) \tag{15.224}$$

Replacing R and I as defined by Equations (15.217) and (15.218)

$$\tan\left(\phi\right) = \frac{B\tan\left(-\theta\omega\right) - A}{A\tan\left(-\theta\omega\right) + B} \tag{15.225}$$

The stability limit is defined as the point where the process oscillation is sustained at constant amplitude. The amplitude ratio is therefore 1. The phase lag between the controller output and the process input will then be π radians (180^{O}). If ϕ is π, then tan(ϕ) will be zero. Therefore, from Equation (15.225)

$$\tan\left(-\theta\omega\right) = \frac{A}{B} \qquad \text{or} \qquad \tan\left(\theta\omega\right) = -\frac{A}{B} \tag{15.226}$$

Replacing A and B as defined by Equations (15.219) and (15.220)

$$\tan\left(\theta\omega\right) = -\frac{T_d\omega - \dfrac{1}{T_i\omega} - \tau\omega}{\dfrac{\tau}{T_i} - \tau T_d\omega^2 - 1} \tag{15.227}$$

This equation may be solved (iteratively or graphically) for ω to give the frequency of the sustained oscillation. If AR is 1, then from Equation (15.222)

$$K_p = \frac{\tau^2\omega^2 + 1}{K_c\sqrt{A^2 + B^2}} \tag{15.228}$$

Replacing A and B using Equations (15.219) and (15.220) gives

$$K_p = \frac{\sqrt{\tau^2\omega^2 + 1}}{K_c\sqrt{\left(T_d\omega - \frac{1}{T_i\omega}\right)^2 + 1}} \tag{15.229}$$

So, once ω is known, we can determine K_p. Should the process gain increase beyond this value, the controller will become unstable.

To plot the stability envelope, as shown in Figures 3.66 and 3.67, we vary θ and, for each value, determine ω and hence the maximum value of K_p. The maximum value of θ, beyond which the controller becomes unstable, is that for which the maximum K_p is equal to the actual K_p.

Taking the same approach with an integrating process

$$G_p(s) = \frac{K_p e^{-\theta s}}{s} \tag{15.230}$$

$$G_p(j\omega) = \frac{K_p e^{-\theta j\omega}}{j\omega} \tag{15.231}$$

Substituting from Equation (15.211) and multiplying both numerator and denominator by j gives

$$G_p(j\omega) = \frac{K_p}{\omega}\left[\sin(-\theta\omega) - j.\cos(-\theta\omega)\right] \tag{15.232}$$

As before, using the same PID controller as Equation (15.213), where now

$$R = \frac{K_p K_c}{\omega}\left[A.\sin(-\theta\omega) + B.\cos(-\theta\omega)\right] \tag{15.233}$$

$$I = \frac{K_p K_c}{\omega}\left[B.\sin(-\theta\omega) - A.\cos(-\theta\omega)\right] \tag{15.234}$$

$$A = 1 \qquad \text{and} \qquad B = -\left(T_d\omega - \frac{1}{T_i\omega}\right) \tag{15.235}$$

From Equation (15.226)

$$\tan(\theta\omega) = \frac{1}{T_d\omega - \frac{1}{T_i\omega}} \tag{15.236}$$

If AR is 1, then from Equations (15.206), (15.233) and (15.234)

$$K_p = \frac{\omega}{K_c\sqrt{A^2 + B^2}} \tag{15.237}$$

Replacing A and B, using Equation (15.235), gives

$$K_p = \frac{\omega}{K_c\sqrt{1 + \left(T_d\omega - \frac{1}{T_i\omega}\right)^2}} \tag{15.238}$$

Equations (15.236) and (15.238) can be used to construct the stability envelope using the same iterative method as described for the self-regulating process. For example, if we consider a PI controller, then Equation (15.236) becomes

$$\tan(\theta\omega) = -T_i\omega \tag{15.239}$$

Substituting this into Equation (15.238), with T_d set to zero, gives

$$K_p.K_c = \omega.\sin(\theta\omega) \tag{15.240}$$

Solving Equation (15.239) for a range of θ and T_i and using the result in equation (15.240) enables the stability envelope shown in Figure 3.68 to be plotted.

Since the oscillation does not require a change in SP to trigger it, the approach described above will work equally well whether control action is based on error or PV. However, we have assumed analog control and so, for digital control, the prediction will be unreliable if process dynamics are of a similar order of magnitude to scan interval. If the technique is applied to other versions of the PID algorithm, for example those that include a derivative filter, then Equation (15.213) should be modified to reflect this.

15.12 Ziegler-Nichols Tuning from Process Dynamics

The Ziegler-Nichols closed loop tuning method involves adjusting the gain of a proportional-only controller until a sustained oscillation is achieved. For such a controller T_i is infinite and T_d zero. Substituting these values into Equation (15.227) gives

$$\tan(\theta\omega) = -\tau\omega \tag{15.241}$$

The period of oscillation (P_u) is the reciprocal of the frequency and so

$$\omega = \frac{2\pi}{P_u} \tag{15.242}$$

Substituting in Equation (15.241) gives Equation (3.94) – reproduced here

$$\frac{2\pi\tau}{P_u} + \tan\left(\frac{2\pi\theta}{P_u}\right) = 0 \tag{15.243}$$

P_u and τ must be positive, and so from Equation (15.243)

$$\tan\left(\frac{2\pi\theta}{P_u}\right) < 0 \tag{15.244}$$

$$\therefore \frac{\pi}{2} < \frac{2\pi\theta}{P_u} < \pi \qquad \text{or} \qquad 2\theta < P_u < 4\theta \tag{15.245}$$

Similarly, to obtain K_u for a proportional-only controller, Equation (15.229) becomes

$$K_p = \frac{\sqrt{\tau^2\omega^2 + 1}}{K_u} \tag{15.246}$$

Combining Equations (15.221) and (15.224)

$$\sqrt{\tan^2(\theta\omega) + 1} = \pm\frac{1}{\cos^2(\theta\omega)} \tag{15.247}$$

Combining Equation (15.241), Equation (15.246) and Equation (15.247) gives

$$K_u = \frac{\pm 1}{K_p \cos(\theta\omega)} \tag{15.248}$$

In order to satisfy the constraint identified by Equation (15.245) we must select the negative root, and so we obtain Equation (3.93) – reproduced here

$$K_u = \frac{-1}{K_p \cos\left(\frac{2\pi\theta}{P_u}\right)} \tag{15.249}$$

We saw, in Chapter 3, that analog tuning methods that do not take account of MV overshoot suggest that K_c (and hence K_u) should be infinite if θ is zero. Equation (15.249) would appear to contradict this. However, from Equation (15.245), P_u will also be zero and thus Equation (15.249) is indeterminate.

For a proportional-only controller on an integrating process the derivation of the Ziegler-Nichols parameters requires a slightly modified approach. From Equation (15.236), without the integral and derivative terms, $\tan(\theta\omega)$ becomes infinite and so

$$\theta\omega = \frac{\pi}{2} \tag{15.250}$$

Combining with Equation (15.242) gives Equation (3.98) – reproduced here

$$P_u = \frac{2\pi}{\omega} = 4\theta \tag{15.251}$$

This result is confirmed by equating T_i from Equations (3.101) and (3.108) and also by equating T_i and T_d from Equations (3.102) and (3.109). If the Ziegler-Nichols closed and open loop methods are to give

consistent values for T_i and T_d, then P_u must be 4θ. This suggests that the process simulated by Ziegler and Nichols might be of the integrating type. A similar condition can be derived for all of their P, PI and PID controller tuning equations to give consistent values for K_c.

$$K_u = \frac{2\tau}{K_p\theta} \tag{15.252}$$

For sustained oscillation of a proportional-only controller Equation (15.238) becomes

$$K_p = \frac{\omega}{K_u} \tag{15.253}$$

Substituting from Equation (15.251) gives Equation (3.97), reproduced here as Equation (15.254), which is inconsistent with Equation (15.252). This illustrates that the two Ziegler-Nichols methods cannot, even for an integrating process, give the same value for K_c – except under the very unlikely circumstance that τ is $\pi/4$.

$$K_u = \frac{\pi}{2K_p\theta} \tag{15.254}$$

This is, however, consistent with the expectation that K_c (a multiple of K_u) would be infinite for a process with no deadtime.

Using the trigonometric identity, as used in Equation (15.247), Equation (15.249) can be rearranged as

$$K_p = \frac{\sqrt{\tan^2\left(\frac{2\pi\theta}{P_u}\right)+1}}{K_u} \tag{15.255}$$

Rearranging Equation (15.243) and multiplying by θ gives:

$$\frac{\theta}{\tau} = \frac{-2\pi\left(\frac{\theta}{P_u}\right)}{\tan\left(2\pi\left(\frac{\theta}{P_u}\right)\right)} \tag{15.256}$$

These equations would appear to suggest that conventional process dynamics, in the form used by most PID tuning methods, can be derived from K_u and P_u. However, while we know P_u, we do not know θ and hence θ/P_u. We showed that P_u must be between 2θ and 4θ and so θ/P_u must be between 0.25 and 0.5. Figures 15.27 and 15.28 show just how much the process dynamics vary with the value of this parameter – again showing that the Ziegler-Nichols closed loop method of determining tuning constants does not uniquely define the process dynamics of a self-regulating process. However, rearranging Equations (15.254) and (15.251) does permit derivation of the dynamics of an integrating process – provided θ is not zero.

$$K_p = \frac{2\pi}{K_u P_u} \tag{15.257}$$

$$\theta = \frac{P_u}{4} \tag{15.258}$$

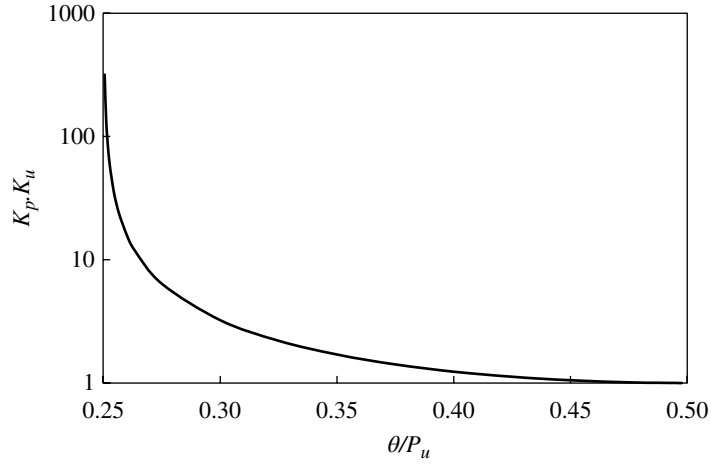

Figure 15.27 Derivation of K_p from K_u and P_u

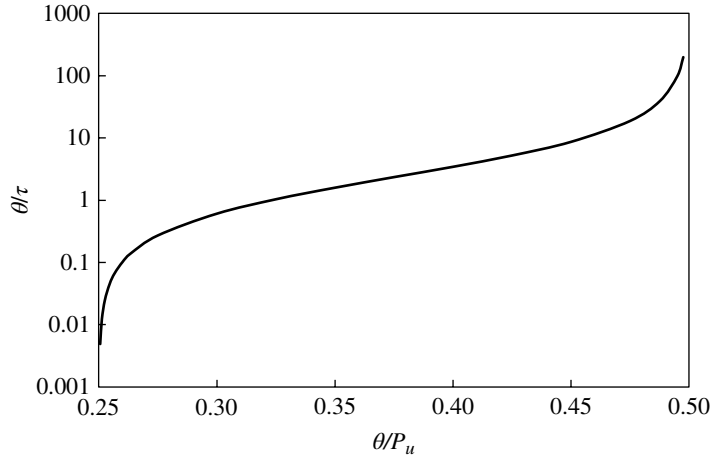

Figure 15.28 Derivation of θ/τ from K_u and P_u

15.13 Partial Fractions

The method of *partial fractions* is most often used by control engineers in developing z-transforms, which in turn are used to convert analog systems to digital. It is best understood by considering a number of examples. The function $f(x)$ can be expanded into the sum of two partial fractions

$$f(x) = \frac{x+1}{(x+2)(x+3)} = \frac{\alpha_1}{x+2} + \frac{\alpha_2}{x+3} \tag{15.259}$$

If we multiply both sides of the equation by $(x + 2)(x + 3)$ we obtain

$$x+1=\alpha_1\left(x+3\right)+\alpha_2\left(x+2\right)=\left(\alpha_1+\alpha_2\right)x+3\alpha_1+2\alpha_2 \tag{15.260}$$

By equating the coefficients of x

$$1=\alpha_1+\alpha_2 \tag{15.261}$$

By equating the constants

$$1=3\alpha_1+2\alpha_2 \tag{15.262}$$

Solving Equations (15.261) and (15.262) for α_1 and α_2 and then substituting the result into Equation (15.259)

$$\frac{x+1}{\left(x+2\right)\left(x+3\right)}=\frac{-1}{x+2}+\frac{2}{x+3} \tag{15.263}$$

There are a number of alternative methods which can often be applied more quickly. The most common of these is the *Heaviside expansion*. This involves first multiplying the function by one of the factors in the denominator and then choosing a value of x that makes this factor zero. For example, if we multiply Equation (15.259) by $(x + 2)$ we obtain

$$\frac{x+1}{x+3}=\alpha_1+\frac{x+2}{x+3}\alpha_2 \tag{15.264}$$

Setting x equal to –2 gives us the result that α_1 is –1. Similarly if we multiply Equation (15.259) by $(x + 3)$ we obtain

$$\frac{x+1}{x+2}=\frac{x+3}{x+2}\alpha_1+\alpha_2 \tag{15.265}$$

Setting x to –3 gives the result that α_2 is 2.

If factors are repeated in the denominator we must include a partial fraction for each of the powers of that factor. For example the factor x^2 'hides' the factor x and so both must be included in the next example.

$$\frac{x+1}{x^2\left(x+2\right)}=\frac{\alpha_1}{x}+\frac{\alpha_2}{x^2}+\frac{\alpha_3}{x+2} \tag{15.266}$$

As always, we can identify α_1, α_2 and α_3 by first multiplying the equation by all the factors.

$$x+1=\alpha_1 x\left(x+2\right)+\alpha_2\left(x+2\right)+\alpha_3 x^2 \tag{15.267}$$

By equating the coefficients of each of the powers of x we obtain

$$1=2\alpha_2 \tag{15.268}$$

$$1=2\alpha_1+\alpha_2 \tag{15.269}$$

$$0=\alpha_1+\alpha_3 \tag{15.270}$$

Solving these gives

$$\alpha_1 = \frac{1}{4} \qquad \alpha_2 = \frac{1}{2} \qquad \alpha_3 = -\frac{1}{4} \tag{15.271}$$

To apply the Heaviside method where there are repeat factors needs a slightly modified approach. If we multiply Equation (15.266) by the factor x we obtain

$$\frac{x+1}{x(x+2)} = \alpha_1 + \frac{\alpha_2}{x} + \frac{\alpha_3 x}{x+2} \tag{15.272}$$

But, because the equation becomes indeterminate, we cannot then find α_1 by setting x equal to zero. However, we can set x to -1 to obtain the equation

$$0 = \alpha_1 - \alpha_2 - \alpha_3 \tag{15.273}$$

We then proceed as normal and multiply Equation (15.266) by the factor x^2.

$$\frac{x+1}{x+2} = \alpha_1 x + \alpha_2 + \frac{\alpha_3 x^2}{x+2} \tag{15.274}$$

Setting x to zero in this equation gives the result we expect for α_2. Similarly, multiplying Equation (15.266) by the factor $(x + 2)$ gives

$$\frac{x+1}{x^2} = \frac{x+2}{x}\alpha_1 + \frac{x+2}{x^2}\alpha_2 + \alpha_3 \tag{15.275}$$

Setting x to -2 gives the result for α_3. We can now solve Equation (15.273) to give the expected result for α_1.

We can also use differentiation to help simplify the equation. If we differentiate Equation (15.267)

$$1 = 2\alpha_1 x + 2\alpha_1 + \alpha_2 + 2\alpha_3 x \tag{15.276}$$

Equating coefficients gives

$$1 = 2\alpha_1 + \alpha_2 \tag{15.277}$$

$$0 = 2\alpha_1 + 2\alpha_3 \tag{15.278}$$

Equation (15.278) could also have been obtained by differentiating Equation (15.276). Solving Equations (15.277) and (15.278) requires either α_2 or α_3 to be derived from the normal Heaviside method.

15.14 *z*-Transforms and Finite Difference Equations

We use z-transforms to describe the behaviour of processes and controllers in much the same way as we use Laplace transforms. Laplace transforms should be strictly only applied to analog systems, while z-transforms are used for discrete systems. Converting a continuous function to its discrete equivalent has two main applications. It allows us to determine process dynamics from discrete process data, i.e. data that has been collected periodically. It also allows us to implement a control algorithm, described in analog form, into a digital system.

As with Laplace transforms, z-transforms can be developed without a detailed understanding of the underlying mathematics. By definition, where s is the Laplace operator and ts is the scan interval,

$$z = e^{s.ts} \qquad \text{and} \qquad \therefore e^{-s.ts} = z^{-1} \tag{15.279}$$

In the same way that others have published tables of transformations of Laplace transforms, they have also provided them for z-transforms. A selection has been included in Table 15.3. A z-transform generates values at a discrete times. During the interval between these times its output is zero. A *zero order hold* will reconstruct the analog signal by holding the value for the duration of the interval – so generating a *staircase* output. It is possible to employ higher order holds. For example, a *first order hold* will linearly extrapolate from the last two outputs and so ramp the output during the interval. This tends to introduce spikes as the analog signal changes direction. A *delayed first order hold* will interpolate between the current value and the next, more closely matching the analog signal but delayed by one scan interval.

The zero order hold (*ZOH*) is most commonly used. It is constructed from two steps, the second in the opposite direction and delayed by the scan interval ts. From Equations (15.90) and (15.91)

$$ZOH = \frac{1}{s} - \frac{e^{-s.ts}}{s} = \frac{1 - e^{-s.ts}}{s} \tag{15.280}$$

For example, consider the first order lag given by Equation (15.105). The zero order hold function is applied before conversion to a z-transform. To proceed we first split the Laplace transform into partial fractions, using the method described in the previous section.

$$\frac{PV}{MV} = \frac{1 - e^{-s.ts}}{s} \frac{K_p}{\tau s + 1} = \frac{K_p\left(1 - e^{-s.ts}\right)}{s} - \frac{K_p\left(1 - e^{-s.ts}\right)\tau}{\tau s + 1} \tag{15.281}$$

And so

$$\frac{PV}{MV} = K_p\left(1 - e^{-s.ts}\right)\left[\frac{1}{s} - \frac{\tau}{\tau s + 1}\right] \tag{15.282}$$

Using Table 15.3 to convert each term to its z-transform

$$\frac{PV}{MV} = K_p\left(1 - z^{-1}\right)\left[\frac{1}{1 - z^{-1}} - \frac{1}{1 - e^{-ts/\tau} z^{-1}}\right] = \frac{K_p\left(1 - e^{-ts/\tau}\right)z^{-1}}{1 - e^{-ts/\tau} z^{-1}} \tag{15.283}$$

Rearranging

$$PV - e^{-ts/\tau} z^{-1} PV = K_p\left(1 - e^{-ts/\tau}\right) z^{-1} MV \tag{15.284}$$

We can convert this to a finite difference equation. The transform z^0 applied to PV is the current value of PV, i.e. PV_n. Applying the transform z^{-1} gives PV_{n-1}, etc. Similarly applying the transform z^{-1} to MV gives MV_{n-1} etc. So we can rewrite Equation (15.284) as

$$PV_n - e^{-ts/\tau} PV_{n-1} = K_p\left(1 - e^{-ts/\tau}\right) MV_{n-1} \tag{15.285}$$

Or

$$PV_n = e^{-ts/\tau} PV_{n-1} + K_p\left(1 - e^{-ts/\tau}\right) MV_{n-1} \tag{15.286}$$

Usually this equation is modified to use MV_n rather than MV_{n-1}.

To include deadtime we modify the z-transforms that are applied to the input. For example z^0 applied to MV becomes $z^{-\theta/ts}$, z^{-1} becomes $z^{-1-\theta/ts}$, etc. In effect this bases the output on the appropriately older value(s) of MV.

$$PV_n = e^{-ts/\tau} PV_{n-1} + K_p \left(1 - e^{-ts/\tau}\right) MV_{n-\theta/ts} \tag{15.287}$$

And, to allow for the fact that PV is unlikely to be zero when MV is zero, we include a bias term.

$$PV_n = e^{-ts/\tau} PV_{n-1} + \left(1 - e^{-ts/\tau}\right)\left(K_p MV_{n-\theta/ts} + bias\right) \tag{15.288}$$

This is the equation used to identify FOPDT models from periodic process data. See Equations (2.16) and (2.17).

Adding the zero order hold to the second order model (including inverse response) described by Equation (15.131) gives

$$\frac{PV}{MV} = \frac{1 - e^{-s.ts}}{s} \frac{K_p\left(\tau_3 s + 1\right)}{\left(\tau_1 s + 1\right)\left(\tau_2 s + 1\right)} \tag{15.289}$$

Converting to partial fractions

$$\frac{PV}{MV} = \frac{\alpha_1}{s} + \frac{\alpha_2}{\tau_1 s + 1} + \frac{\alpha_3}{\tau_2 s + 1} \tag{15.290}$$

$$\alpha_1 = K_p\left(1 - e^{-s.ts}\right) \tag{15.291}$$

$$\alpha_2 = K_p\left(1 - e^{-s.ts}\right) \frac{\tau_1\left(\tau_3 - \tau_1\right)}{\tau_1 - \tau_2} \tag{15.292}$$

$$\alpha_3 = K_p\left(1 - e^{-s.ts}\right) \frac{\tau_2\left(\tau_2 - \tau_3\right)}{\tau_1 - \tau_2} \tag{15.293}$$

Using Table 15.3 to convert to z-transform

$$\frac{PV}{MV} = K_p \left[1 + \frac{\tau_1\left(\tau_3 - \tau_1\right)}{\tau_1 - \tau_2} \frac{1 - z^{-1}}{\left(1 - e^{-ts/\tau_1} z^{-1}\right)\tau_1} + \frac{\tau_2\left(\tau_2 - \tau_3\right)}{\tau_1 - \tau_2} \frac{1 - z^{-1}}{\left(1 - e^{-ts/\tau_2} z^{-1}\right)\tau_2}\right] \tag{15.294}$$

Rearranging

$$\frac{PV}{MV} = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}} \tag{15.295}$$

$$a_1 = e^{-ts/\tau_1} + e^{-ts/\tau_2} \tag{15.296}$$

$$a_2 = -e^{-ts/\tau_1} e^{-ts/\tau_2} \tag{15.297}$$

$$b_1 = K_p \left[1 + \frac{\tau_3 - \tau_1}{\tau_1 - \tau_2} e^{-ts/\tau_1} + \frac{\tau_2 - \tau_3}{\tau_1 - \tau_2} e^{-ts/\tau_2}\right] \tag{15.298}$$

$$b_2 = K_p\left[e^{-ts/\tau_1}e^{-ts/\tau_2} + \frac{\tau_2-\tau_3}{\tau_1-\tau_2}e^{-ts/\tau_1} + \frac{\tau_3-\tau_1}{\tau_1-\tau_2}e^{-ts/\tau_2}\right] \tag{15.299}$$

Converting to finite difference form enables second order processes with inverse response to be modelled from periodic process data. If $\tau_3 = 0$, then making this substitution into Equations (15.298) and (15.299) will give revised coefficients for a simple second order process. However, if $\tau_1 = \tau_2$, we cannot take the same approach since b_1 and b_2 are indeterminate. Instead we have to derive the model from first principles. Taking the usual approach of multiplying by the zero order hold and splitting into partial fractions, we obtain

$$\frac{PV}{MV} = \frac{1-e^{-s.ts}}{s}\frac{K_p(\tau_3 s+1)}{(\tau s+1)^2} = K_p\left(1-e^{-s.ts}\right)\left[\frac{\alpha_1}{s} + \frac{\alpha_2}{\tau s+1} + \frac{\alpha_3}{(\tau s+1)^2}\right] \tag{15.300}$$

$$\alpha_1 = K_p\left(1-e^{-s.ts}\right) \tag{15.301}$$

$$\alpha_2 = -K_p\left(1-e^{-s.ts}\right)\tau \tag{15.302}$$

$$\alpha_3 = K_p\left(1-e^{-s.ts}\right)(\tau_3-\tau) \tag{15.303}$$

Converting to z-transform

$$\frac{PV}{MV} = K_p\left(1-z^{-1}\right)\left[\frac{1}{1-z^{-1}} - \frac{1}{1-e^{-ts/\tau}z^{-1}} + \frac{\tau_3-\tau}{\tau^2}\frac{ts.e^{-ts/\tau}z^{-1}}{\left(1-e^{-ts/\tau}z^{-1}\right)^2}\right] \tag{15.304}$$

Converting to finite difference form gives the same form as Equation (15.295) but with the coefficients modified.

$$a_1 = 2e^{-ts/\tau} \tag{15.305}$$

$$a_2 = -e^{-2ts/\tau} \tag{15.306}$$

$$b_1 = K_p\left[1-e^{-ts/\tau} + \frac{(\tau_3-\tau)ts}{\tau^2}e^{-ts/\tau}\right] \tag{15.307}$$

$$b_2 = K_p\left[e^{-2ts/\tau} - e^{-ts/\tau} - \frac{(\tau_3-\tau)ts}{\tau^2}e^{-ts/\tau}\right] \tag{15.308}$$

We can also develop finite difference formulae for control algorithms. Indeed, this is how algorithms are implemented in the DCS. For example, Equation (15.286) also gives the form of the first order exponential filter. If we set K_p to 1, replace τ with τ_f, PV with Y and MV with X we get

$$Y_n = e^{-ts/\tau_f}Y_{n-1} + \left(1-e^{-ts/\tau_f}\right)X_{n-1} \tag{15.309}$$

Comparison with Equations (5.29) and (5.30) shows that the filter uses X_n rather than X_{n-1}. This is the convention adopted by most DCS vendors. In theory it will perform slightly differently but the change in most

cases will be unnoticeable. An approximate form of the first order exponential filter can be developed from applying the Taylor approximation to Equation (15.279).

$$z^{-1} = e^{-s.ts} \approx 1 - s.ts \qquad \therefore s = \frac{1 - z^{-1}}{ts} \tag{15.310}$$

Replacing s in the Laplace form of the filter

$$\frac{Y}{X} = \frac{1}{\tau_f s + 1} \approx \frac{ts}{\tau_f \left(1 - z^{-1}\right) + ts} \tag{15.311}$$

Rearranging

$$Y = \frac{\tau_f z^{-1} Y + ts.X}{\left(\tau_f + ts\right)} \tag{15.312}$$

Converting to finite difference form

$$Y_n = \frac{\tau_f}{\tau_f + ts} Y_{n-1} + \frac{ts}{\tau_f + ts} X_n \tag{15.313}$$

This the same as the form described by Equations (5.29) and (5.33).

We can develop a second order filter by making the substitution $\tau_3 = 0$ in Equations (15.295), (15.305), (15.306), (15.307) and (15.308). We also set K_p to 1, replace PV with Y, replace MV with X and define P as $e^{-ts/\tau}$.

$$Y_n = 2P.Y_{n-1} - P^2 Y_{n-2} + \left(1 - P + P\ln\left(P\right)\right)X_n - P\left(1 - P + \ln\left(P\right)\right)X_{n-1} \tag{15.314}$$

Due to the different method of deriving the formula the result is considerably different from that shown by setting $P_1 = P_2 = P$ in Equation (15.22) which was developed by combining two first order filters. However, the difference in performance is unlikely to be perceptible.

The finite difference form of the PID algorithm can be developed from applying the same approximation; replacing s in Equation (15.135) with that from Equation (15.310) gives

$$\frac{M}{E} = K_c\left[1 + \frac{1}{T_i s} + T_d s\right] = K_c\left[1 + \frac{ts}{T_i\left(1 - z^{-1}\right)} + T_d \frac{1 - z^{-1}}{ts}\right] \tag{15.315}$$

Multiplying by $(1 - z^{-1})$ and rearranging

$$\left(1 - z^{-1}\right)M = K_c\left[\left(1 - z^{-1}\right) + \frac{ts}{T_i} + \frac{T_d}{ts}\left(1 - 2z^{-1} + z^{-2}\right)\right]E \tag{15.316}$$

Converting to finite difference form

$$M_n - M_{n-1} = K_c\left[\left(E_n - E_{n-1}\right) + \frac{ts}{T_i}E_n + \frac{T_d}{ts}\left(E_n - 2E_{n-1} + E_{n-2}\right)\right] \tag{15.317}$$

This is the equation we developed by approximating continuous functions to give Equation (3.33). It is possible also to apply the Padé approximation to Equation (15.279).

$$z^{-1} = e^{-s.ts} \approx \frac{2 - s.ts}{2 + s.ts} \qquad \text{or} \qquad s \approx \frac{2}{ts}\frac{1 - z^{-1}}{1 + z^{-1}} \tag{15.318}$$

This is known as *Tustin's method* and, without showing the derivation, results in a different form of the PID algorithm.

$$M_n - M_{n-2} = K_c\left[\left(E_n - E_{n-2}\right) + \frac{ts}{T_i}\left(\frac{E_n + 2E_{n-1} + E_{n-2}}{2}\right) + \frac{2T_d}{ts}\left(E_n - 2E_{n-1} + E_{n-2}\right)\right] \tag{15.319}$$

The method of determining M based not on the previous scan, but the one previous to that, might at first seem rather odd. Indeed, this view would appear to be that of the control system vendors since none have used it. However, as a closer approximation to the analog algorithm, there is no reason why it should not work. Inspection of each of the P and I parts of the algorithm shows that each could be represented by the sum of two consecutive moves. While not exactly the same as two consecutive moves the D action is doubled. We can therefore write an algorithm that, in two steps, would make almost the same correction.

$$M_n - M_{n-1} = K_c\left[\left(E_n - E_{n-1}\right) + \frac{ts}{T_i}\left(\frac{E_n + E_{n-1}}{2}\right) + \frac{T_d}{ts}\left(E_n - 2E_{n-1} + E_{n-2}\right)\right] \tag{15.320}$$

This is the algorithm developed as Equation (3.48) by applying the trapezium rule.

The lead-lag/deadtime algorithm is described by the Laplace transform

$$\frac{Y}{X} = Ke^{-\theta s}\frac{T1.s + 1}{T2.s + 1} \tag{15.321}$$

Assume for the moment that $\theta = 0$. Multiplying by the zero order hold and splitting into partial fractions gives

$$\frac{Y}{X} = K\frac{1 - e^{-s.ts}}{s}\frac{T1.s + 1}{T2.s + 1} = K\left(1 - e^{-s.ts}\right)\left[\frac{1}{s} + \frac{T1 - T2}{T2.s + 1}\right] \tag{15.322}$$

Converting to z-transform

$$\frac{Y}{X} = K\left(1 - z^{-1}\right)\left[\frac{1}{1 - z^{-1}} + \frac{T1 - T2}{\left(1 - e^{-ts/T2}z^{-1}\right)T2}\right] \tag{15.323}$$

Rearranging

$$\frac{Y}{X} = \frac{K\left[\dfrac{T1}{T2} - \left(\dfrac{T1 - T2}{T2} + e^{-ts/T2}\right)z^{-1}\right]}{1 - e^{-ts/T2}z^{-1}} \tag{15.324}$$

Converting to finite difference form

$$Y_n = e^{-ts/T2}Y_{n-1} + K\frac{T1}{T2}X_n - K\left(\frac{T1 - T2}{T2} + e^{-ts/T2}\right)X_{n-1} \tag{15.325}$$

Including deadtime gives the more usual discrete version of the algorithm as described by Equation (6.12).

References

1. Ziegler J.G. and Nichols N.B. (1942) Optimum settings for automatic controllers. *Transactions of the ASME*, **64**, 759–768.
2. Sundaresan R. and Krishnaswamy P.R. (1977) Estimation of time delay, time constant parameters in time, frequency and Laplace domains. *Canadian Journal of Chemical Engineering*, **56**, 257–262.
3. Smith C.L. (1972) *Digital Computer Process Control*. Intext Educational Publishers, 147.
4. Process Instrumentation Terminology. The Instrumentation, Systems and Automation Society, ISA-51.1-1979 (R1993).
5. O'Dwyer A. (2000) A summary of PI and PID controller tuning rules for processes with time delay. IFAC Digital Control: Past, Present and Future of PID Control, Terrassa, Spain.
6. Lopez A.M., Miller J.A., Smith C.L. and Murrill P.W. (1967) Tuning controllers with error-integral criteria, *Instrumentation Technology*, **14**, 57–62.
7. Åström K. and Hägglund T. (1984) Automatic tuning of simple regulators with specifications on phase and amplitude margins. *Automatica*, **20**, 645–651.
8. Harriott P. (1964) *Process Control*. McGraw-Hill, New York, **180**.
9. Chien K.L., Hrones J.A. and Reswick J.B. (1952) On the automatic control of generalized passive systems. *Transactions of the ASME*, **74**, 175–185.
10. Cohen G.H. and Coon G.A. (1953) Theoretical considerations of retarded control. *Transactions of the ASME*, **75**, 827–834.
11. Smith C.L. (1972) *Digital Computer Process Control*. Intext Educational Publishers, 176.
12. Murrill P.W. (1967) *Automatic Control of Processes*. International Textbook Company, 356.
13. Kaya A. and Scheib T.A. (1988) Tuning of PID controls of different structures. *Control Engineering*, **35**(7), 62–65.
14. Smuts J.F. (2011) *Process Control for Practitioners*. Opticontrol Inc., 52–53.
15. Dahlin E.B. (1968) Designing and tuning digital controllers. *Instruments and Control Systems*, **2**(6), 77–83.
16. Chien I-L. (1988) IMC-PID controller design – an extension. *Proceedings of the IFAC Adaptive Control of Chemical Processes Conference*, Denmark, 147–152.
17. Chien I-L. and Fruehauf P.S. (1990) Consider IMC tuning to improve controller performance. *Chemical Engineering Progress*, **86**, 33–41.
18. Rice R. and Cooper D.J. (2002) Design and tuning of PID controllers for integrating (non-self-regulating) processes. *Proceedings of the ISA Annual Meeting*.
19. Marlin T.E. (2000) *Process Control*, 2nd edition. McGraw-Hill, New York, 278–284.
20. Li J. (2010) A PID tuning method using MINLP with nonparametric process and disturbance models. AIChE Spring Meeting, San Antonio, Texas, paper 96D.
21. Bartman R.V. (1981) Dual composition control in a C_3/C_4 splitter. *Chemical Engineering Progress*, **77**(9), 58–62.
22. Wade H.L. (2004) *Basic and Advanced Regulatory Control: System Design and Application*. ISA, 150–159.
23. Smith O.J.M. (1957) Closer control of loops with deadtime. *Chemical Engineering Progress*, **53**(5), 217–219.
24. Dahlin E.B. (1968) Designing and tuning digital controllers. *Instruments and Control Systems*, **2**(6), 77–83.

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control

25. Bristol E.H. (1966) On a new measure of interactions for multivariable process control. *IEEE Transactions of Automatic Control*, **11**, 133.
26. Yang J., Peng W., Ward M.O. and Rundensteiner E.A. (2003) Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. Worcester Polytechnic Institute.
27. Clevett K.J. (1999) *Process Analyzer Technology*. John Wiley and Sons, Inc., Hoboken, NJ.
28. O'Connell H.E. (1946) Plate efficiency of fractionating columns and absorbers. *Transactions of the AIChE*, **42**, 741.
29. Ryskamp C.J. (1980) New strategy improves dual composition control. *Hydrocarbon Processing*, **59**(6), 51–59.
30. Prediction of vapour pressure of pure hydrocarbons and narrow-boiling petroleum fractions. *API Technical Data Book*, procedure 5A1.19.
31. Jafarey A., Douglas J.M. and McAvoy T.J. (1979) Short-cut techniques for distillation column design and control. *Industrial Engineering Chemistry, Process Design and Development*, **18**(2), 197.
32. Colburn A.P. (1941) The simplified calculation of diffusional processes – II. Number of transfer units or theoretical plates. *Industrial Engineering Chemistry*, **33**, 459–467.
33. Shinskey F.G. (1977) *Distillation Control*, McGraw-Hill Book Company, New York, 227–231 and 306–310.
34. Weber R. and Gaitonde N.Y. (1982) Non-interactive distillation tower analyzer control. *Proceedings of American Control Conference*, Arlington, VA, 87.
35. Smith C.L. (2012) *Distillation Control: An Engineering Perspective*. John Wiley and Sons, Inc., Hoboken, NJ, 178.
36. Martin G.D., Turpin L.E. and Cline R.P. (1991) Estimating control function benefits. *Hydrocarbon Processing*, **70**(6), 68–73.
37. Martin G.D. (2004) Understand control benefits estimates. *Hydrocarbon Processing*, **83**(10), 43–46.
38. Harris T.J. (1989) Assessment of control loop performance. *Canadian Journal of Chemical Engineering*, **67**, 856–861.
39. Thornhill N.F., Huang B. and Zhang H. (2003) Detection of multiple oscillations in control loops. *Journal of Process Control*, **13**, 91–100.
40. Hägglund T. (1995) A control-loop performance monitor. *Control Engineering Practice*, **3**, 1543–1551.
41. *Alarm Systems: A Guide to Design, Management and Procurement*. EEMUA Publication No. 191, third edition (2013).
42. Brown M.B. and Forsythe A.B. (1974) *Journal of the American Statistical Association*, **69**, 364–367.
43. Anderson T.W. (1994) *The Statistical Analysis of Time Series*. John Wiley and Sons, Inc., Hoboken, NJ, 186–187.
44. Horwitz W. (2003) *Journal of the Association of Official Analytical Chemists*, **86**, 109.
45. Boyer K.W., Horwitz W. and Albert R. (1985) *Analytical Chemistry*. **57**, 454–459.
46. Smith S.W. (1999) *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, 319–327.
47. Seborg D.E., Edgar T.F. and Mellichamp D.A. (1989) *Process Dynamics and Control*. John Wiley and Sons, Inc., Hoboken, NJ.

Index

Process Control: A Practical Approach, Second Edition. Myke King.

Companion website: www.wiley.com/go/king/process_control